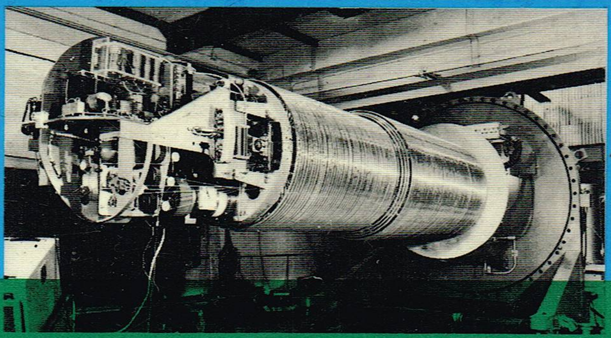
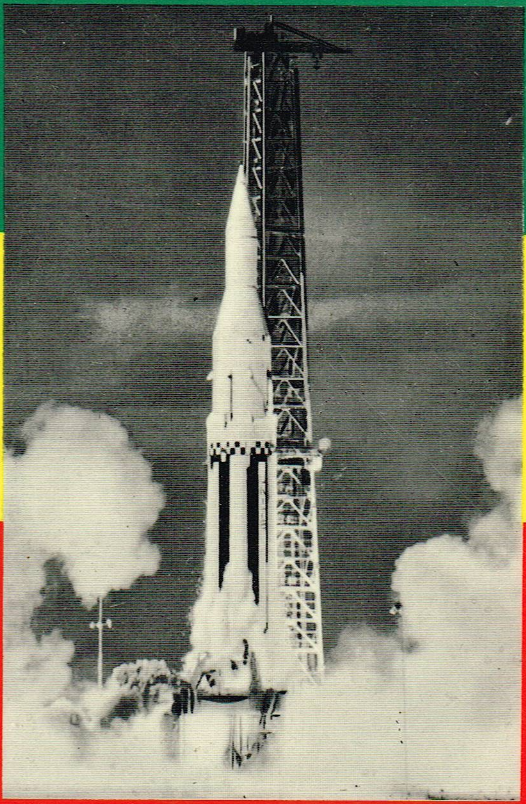
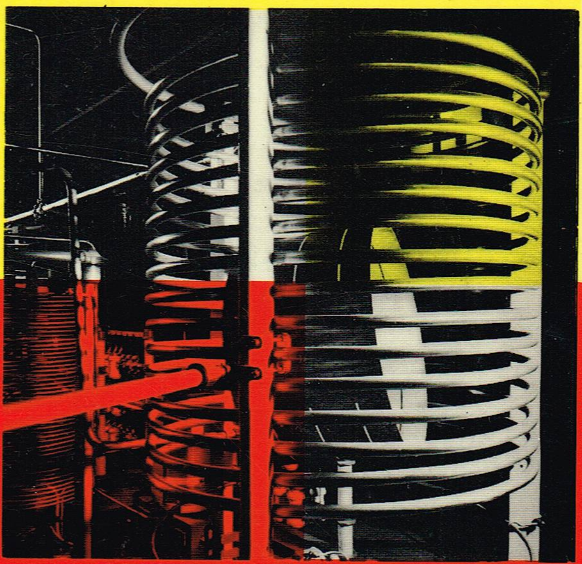
LOUIS LEPRINCE-RINGUET

LA SCIENCE CONTEMPORAINE









les sciences physiques et leurs applications

LAROUSSE

collection in-quarto

LAROUSSE

établis avec la collaboration des meilleurs spécialistes, enrichis d'une illustration à la fois précise et évocatrice, d'une lecture facile et passionnante, les ouvrages in-quarto Larousse forment, par leur ensemble, un magnifique panorama du monde et des activités humaines.

MYTHOLOGIES
2 volumes

LA FRANCE **géographie, tourisme**
2 volumes

LITTÉRATURE FRANÇAISE
2 volumes

HISTOIRE DE FRANCE
2 volumes

HISTOIRE UNIVERSELLE
2 volumes

GÉOGRAPHIE UNIVERSELLE LAROUSSE
3 volumes

ATLAS GÉNÉRAL LAROUSSE

ASTRONOMIE **les astres, l'univers**

L'ART ET L'HOMME
3 volumes

LA TERRE **notre planète**

LA MONTAGNE

LA MER

LA VIE DES PLANTES

LA VIE DES ANIMAUX
2 volumes

LA VIE

LA DÉCOUVERTE DE LA TERRE
en coédition avec PARIS-MATCH

Phot. O.N.E.R.A. - U.S.I.S. - C.S.F. - X.

la science
contemporaine

ouvrage en 2 volumes
publié sous la direction de
Louis Leprince-Ringuet
de l'Académie Française
de l'Académie des Sciences
professeur
au Collège de France

avec la collaboration de
Jean-Louis Rigal
professeur au Centre
universitaire Dauphine

et de

Jacques Clavier
Gabriel de Clavière
Robert Collongues
Maurice Françon
Jean Heidmann
† Dominique Leprince-Ringuet
Georges Parreins
Jacques Parrod
Claude Picard
† Anne-Marie Rozis-Saulgeot
Maurice Vernet-Lozet
pour le premier volume

la science contemporaine

les sciences physiques et leurs applications 1

LIBRAIRIE LAROUSSE

17, rue du Montparnasse, et boulevard Raspail, 114, Paris

table des matières



PRÉSENTATION

par LOUIS LEPRINCE-RINGUET, professeur au Collège de France et à l'École polytechnique.

DEUX HAUTS LIEUX DE LA SCIENCE CONTEMPORAINE

1. Le Centre européen de recherche nucléaire

par LOUIS LEPRINCE-RINGUET, professeur au Collège de France et à l'École polytechnique.



2. L'observatoire radio-astronomique de Nançay

par JEAN HEIDMANN, maître de recherches au C. N. R. S.



JUSQU'AUX FRONTIÈRES DE L'UNIVERS

par † ANNE-MARIE ROZIS-SAULGEOT, professeur à la Faculté des Sciences de Besançon.

LA RADIO-ASTRONOMIE : REGARDS NOUVEAUX SUR L'UNIVERS

par JEAN HEIDMANN, maître de recherches au C. N. R. S.

LES RAYONS COSMIQUES

par LOUIS LEPRINCE-RINGUET, professeur au Collège de France et à l'École polytechnique.

PREMIERS PAS DANS L'ESPACE : FUSÉES ET SATELLITES



par MAURICE VERNET-LOZET, ingénieur de la Société d'étude de la propulsion par réaction,



et par GABRIEL DE CLAVIÈRE, ingénieur civil de l'Aéronautique.







COMMENT SE TRANSMET ET SE DÉGRADE L'INFORMATION

par JACQUES CLAVIER, ingénieur de l'Air.



LES ORDINATEURS AU SERVICE DE L'INTELLIGENCE

par CLAUDE PICARD, directeur de recherches au C. N. R. S. (Institut B. Pascal).



ÉNERGIE ET INDUSTRIE NUCLÉAIRES

par GEORGES PARREINS, ingénieur au Commissariat à l'énergie atomique.



LES HAUTES TEMPÉRATURES

par ROBERT COLLONGUES, professeur à la Faculté des Sciences de Paris.



LES MACROMOLÉCULES : DES MATIÈRES PLASTIQUES A L'A. D. N.

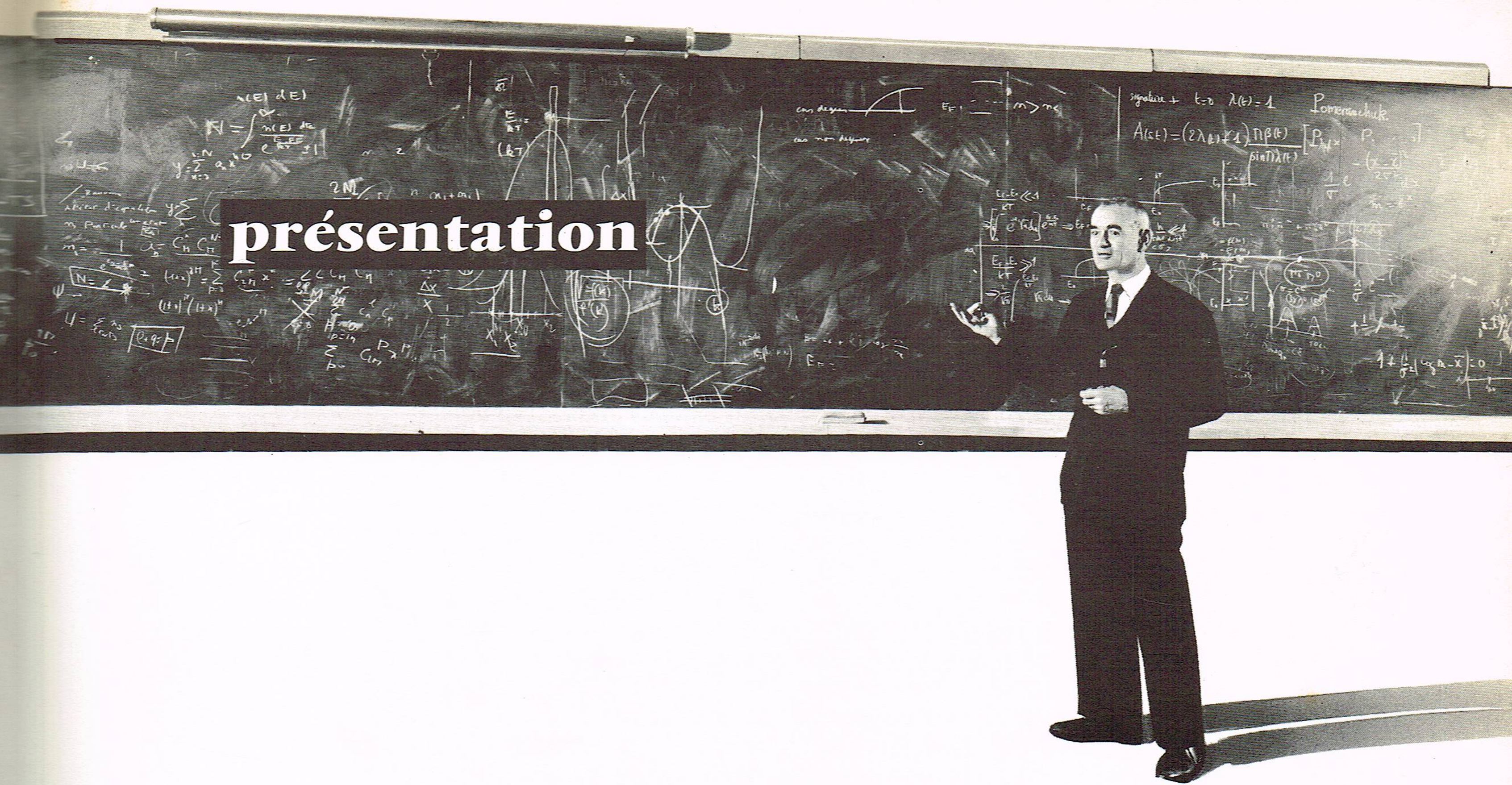
par JACQUES PARROD, professeur à la Faculté des Sciences de Strasbourg.

par M. VEYSSIE-COUNILLON, maître assistant à la Faculté des Sciences de Paris, et G. KUHN, maître de conférences à la Faculté des Sciences de Grenoble.

par MAURICE FRANÇON, professeur à la Sorbonne et à l'Institut d'optique.

L'index général de l'ouvrage se trouve à la fin du tome II.

présentation



On peut envisager de diverses façons un gros ouvrage sur les sciences physiques, et nous souhaitons expliquer ici, en introduction, ce qui a guidé notre choix. Tout d'abord nous savons qu'un nombre constamment croissant d'esprits dont la formation mathématique est faible s'intéresse ardemment aux magnifiques développements scientifiques, et s'interroge : comment pouvoir ne pas rester complètement en arrière, un peu comme des émigrés à côté d'un monde en évolution rapide?

C'est précisément pour ces lecteurs que l'ouvrage est écrit. Notre but est de les informer du progrès scientifique, de leur permettre de comprendre valablement par quel mécanisme une fusée est lancée, quelles sont les lois qui gouvernent le mouvement des satellites artificiels, pourquoi et en utilisant quels phénomènes on a pu donner des ordres à "Mariner II" à plusieurs millions de kilomètres de distance. Mais notre ambition n'est pas limitée au seul terrain de la mécanique et de l'espace. On connaît aussi les données fondamentales des atomes, et également bien des choses sur le noyau même de ces atomes. On en connaît les constituants, et les physiciens jonglent avec des corpuscules bien plus ténus que les atomes, ou même les noyaux, dont ils parviennent à étudier les mœurs — leur création à partir d'énergie, leur désintégration spontanée, leurs interactions — et à comprendre parfois le comportement. On va même jusqu'à créer l'antimatière, cette chose mystérieuse qui correspond à une réalité extrêmement précise et définie.

Ainsi notre objectif final est-il de permettre une compréhension des sciences physiques et de leur évolution rapide. A vrai dire, la science se développe actuellement de façon très complexe et dans des conditions extrêmement variées. Il y a encore, à l'un des pôles, le bureau rempli de livres et de calculs, où le mathématicien passe des journées entières à réfléchir à des problèmes théoriques. Mais il y a également le centre, moins individuel, de physique théorique ou de mathématiques, dans lequel se tiennent plusieurs théoriciens qui travaillent partiellement en commun, discutent ensemble fréquemment et s'entraident finalement, car la plupart des problèmes ont

Phot. Larousse

de multiples faces et exigent une collaboration. Ce genre de centre, d'ailleurs, est de plus en plus répandu ; il se retrouve dans tous les grands ensembles de la physique des particules fondamentales, à Berkeley, à Brookhaven, au CERN ; il en existe de plus en plus dans les grandes universités, comme Orsay, ou dans des ensembles à caractère plus technique comme Saclay ou à l'O.N.E.R.A.

Il y a également le laboratoire expérimental de science de base, dans lequel toute une population, comprenant physiciens, mécaniciens, dessinateurs, chimistes et autres, concourt à une réalisation commune. Dans certains cas, ce sont encore de petits laboratoires avec des équipes d'une dizaine de chercheurs et de techniciens ; ainsi les laboratoires universitaires s'occupant des semi-conducteurs ou de la physique du solide travaillent souvent dans ces conditions. Mais d'autres exigent une complexité plus grande ; quand on examine la façon dont se fait actuellement la découverte en physique des particules fondamentales, on s'aperçoit que presque rien ne peut se développer en dehors de très grands centres puissamment équipés. Un laboratoire ne peut vivre que s'il travaille en liaison avec l'un des trois ou quatre grands centres du monde, et c'est ce qui se passe effectivement. Nous avons à Paris des groupes de chercheurs, et nos instituts de l'Ecole polytechnique, du Collège de France, d'Orsay comprennent un minimum d'une centaine de personnes, toutes attelées au même travail. Des laboratoires comme ceux-ci existent également dans les grandes universités du monde, et ils dépendent étroitement des accélérateurs géants, dont ils tirent fondamentalement leur alimentation scientifique.

Ainsi, dans le domaine de la recherche de base, là où les applications ne sont pas envisagées, d'énormes centres se présentent comme de puissants laboratoires industriels, auxquels ils sont, du point de vue de l'apport technique, tout à fait comparables. Lorsque l'on va au CERN, on trouve des appareils très onéreux, très modernes, très poussés, dont la réalisation constitue souvent une véritable performance industrielle : on ne se sent pas dans un laboratoire de recherche fondamentale classique ; tout s'uniformise en fait, et les interactions entre les industries évoluées et les laboratoires de recherche sont telles que la distinction ne peut plus s'effectuer facilement.

Il y a également des ensembles mixtes, dans lesquels sont acquises et intensément développées les connaissances de base, mais auprès d'une activité plus pratique ; on peut citer, par exemple, Saclay, l'O.N.E.R.A., ou les grands laboratoires des macromolécules, ou encore beaucoup de centres dépendant de groupes industriels. Ce ne sont pas, à vrai dire, des institutions nouvelles, mais c'est leur prodigieux développement qui constitue un phénomène nouveau. Saclay comporte des milliers d'ingénieurs, de chercheurs et de techniciens, qui sont répartis entre des directions s'occupant de la constitution de la matière, du calcul des réacteurs atomiques, de la physique théorique, de la métallurgie, de la biologie, de la chimie radio-active, etc. On ne peut plus dire quelle est la limite entre ce qui est science et ce qui est application ; on peut seulement envisager que, dans un service, certaines applications arrivent en général à court terme, alors que, dans un autre, elles sont plus éloignées. Même chose pour les groupes d'industries. L'ensemble des industries métallurgiques a organisé un grand institut, mais d'autres groupes d'industries ont des laboratoires dans lesquels on fait de la recherche de base, avec publication de thèses de doctorat, et dans lesquels néanmoins on se préoccupe des éventuelles applications à court ou à long terme : ainsi, dans le domaine de l'électronique, les grands laboratoires de la C.S.F. ou le Laboratoire d'électronique et de physique appliquée.

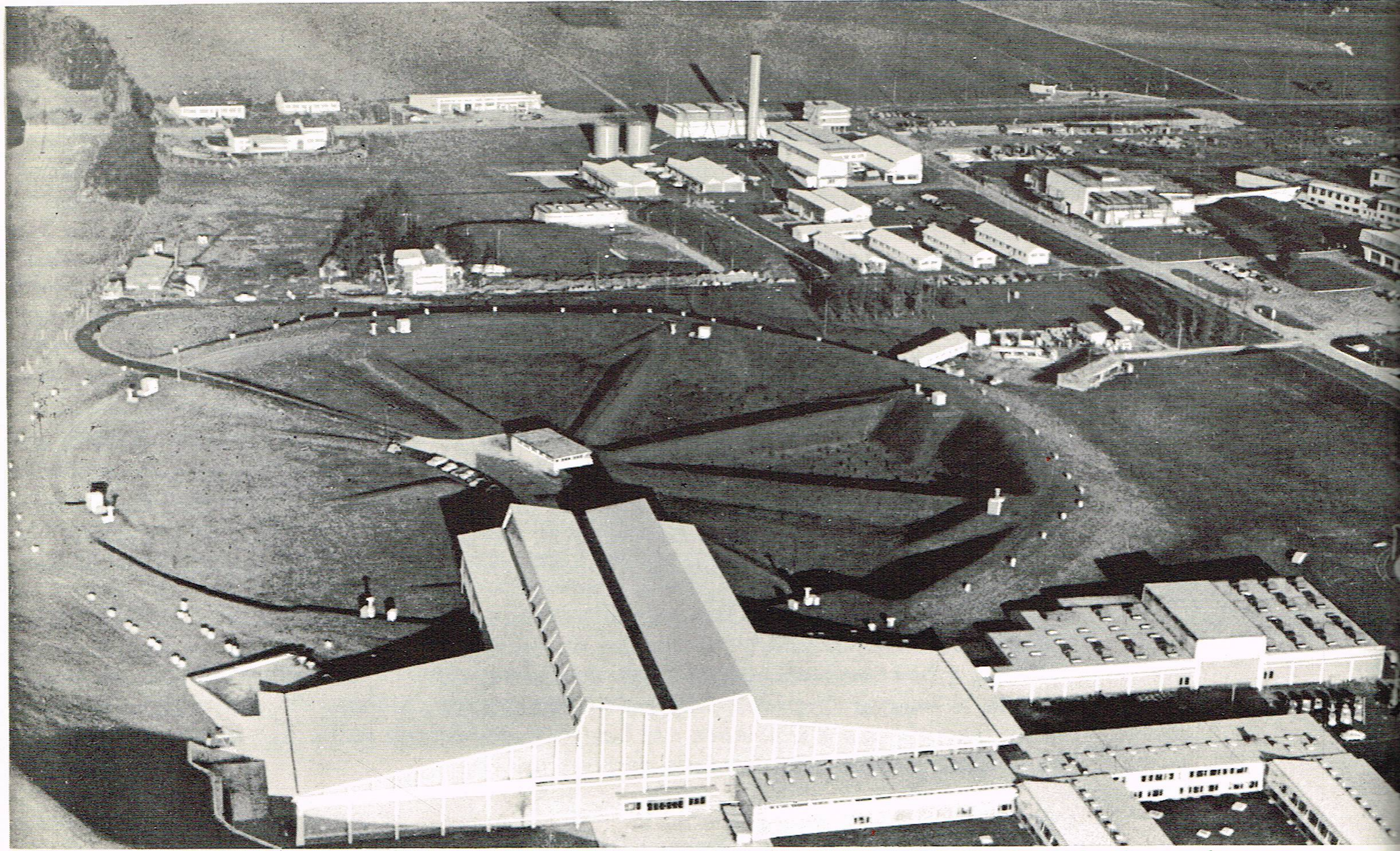
Ainsi va la vie scientifique, avec toutes ses facettes complémentaires. C'est pour mettre en relief cette profonde et vivante liaison entre tous les modes de travail que nous présentons d'abord quelques grands ensembles de l'activité humaine. Nous allons parcourir ainsi quelques-unes des voies impériales de la science, et d'abord celle de l'Espace, si riche d'acquis et de promesses avec l'astronomie, la radio-astronomie, le rayonnement cosmique, et comprenant un grand panorama sur les fusées et les satellites. Ensuite suivra le développement sur un secteur (on pourrait dire un " combinat ") très caractéristique de notre époque, à savoir celui de l'information et de l'électronique, avec les moyens de traiter la première, c'est-à-dire les machines qui s'y rapportent et les ordinateurs, dont l'expansion est extraordinaire, et avec les derniers développements de la seconde, aux réalisations de plus en plus " transistorisées " et " miniaturisées ". Enfin seront traitées quelques grandes disciplines au développement rapide, qui prennent une importance de plus en plus considérable dans notre existence : celle des macromolécules, allant des plastiques jusqu'aux gènes qui transmettent l'hérédité, celle des très hautes températures avec les chalumeaux à plasma, celle de l'optique moderne, avec des possibilités qui étaient inconcevables il y a fort peu d'années, celle de l'énergie nucléaire, dont la puissance, la variété nous étonnent tous les jours.

Mais, avant tout cela, en exergue, pour bien marquer le rythme du travail scientifique, s'impose la présentation de deux grands instituts de recherche de base, deux hauts lieux de la science : le Centre européen de recherche nucléaire (CERN) et le Centre radio-astronomique de Nançay.

Mais, que de problèmes sous-jacents !...
- La synchronisation de l'accélérateur du CERN nous oblige à admettre les résultats de la relativité d'Einstein.
- La diffraction des électrons et les microscopes électroniques nous poussent à comprendre non seulement la physique des ondes, mais également la mécanique quantique, cette grande et superbe constructrion conceptuelle qui explique le monde des ondes et des particules sous leur aspect microscopique.
- Le fonctionnement d'un transistor (plus c'est petit, plus c'est compliqué) implique la compréhension de cette mécanique quantique indispensable et de la physique de l'état solide.

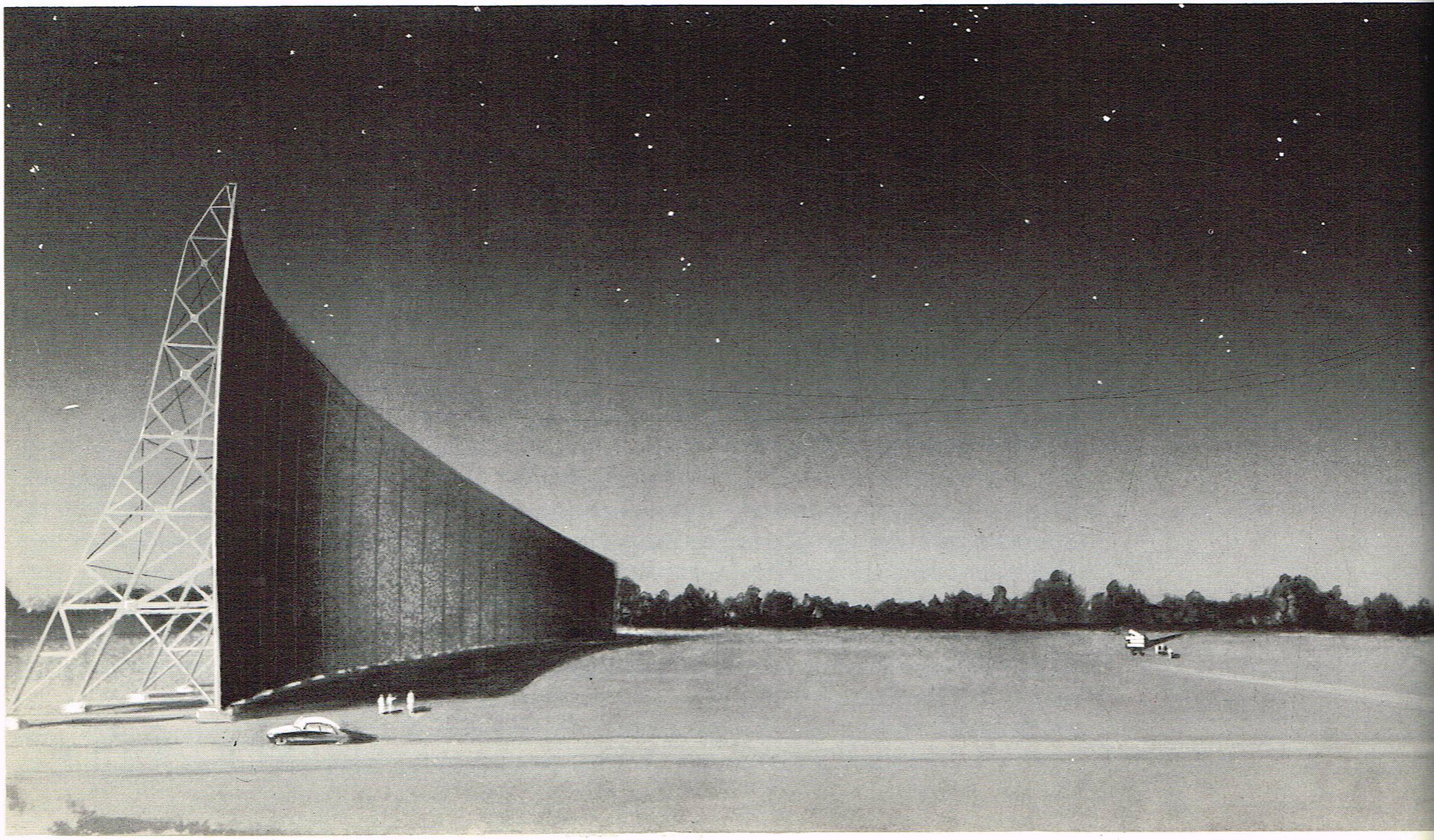
Plus près de notre entendement classique, les problèmes liés aux fusées, aux satellites, à l'exploration de l'espace, ces problèmes si passionnants de rencontres sur orbite, de vols spatiaux vers les planètes, exigent une parfaite compréhension de la mécanique classique.
Mais, après avoir admiré le panorama des développements scientifiques, leur interférence avec les développements des progrès industriels, l'esprit humain doit essayer de comprendre - sinon il ne peut être satisfait. Ce désir d'étancher une légitime soif de connaissance justifiera le tome II, tome de base dans lequel il sera question de la mécanique à l'ère spatiale, de la physique des ondes, de la relativité avec ses curieuses conclusions, de la statistique et des probabilités, du monde agité des particules, de la mécanique quantique, de la connaissance de l'état solide, des bases de la théorie de l'information. Pour exposer cet ensemble, un très grand effort d'adaptation a été fait : nous en reparlerons en le présentant de façon particulière.

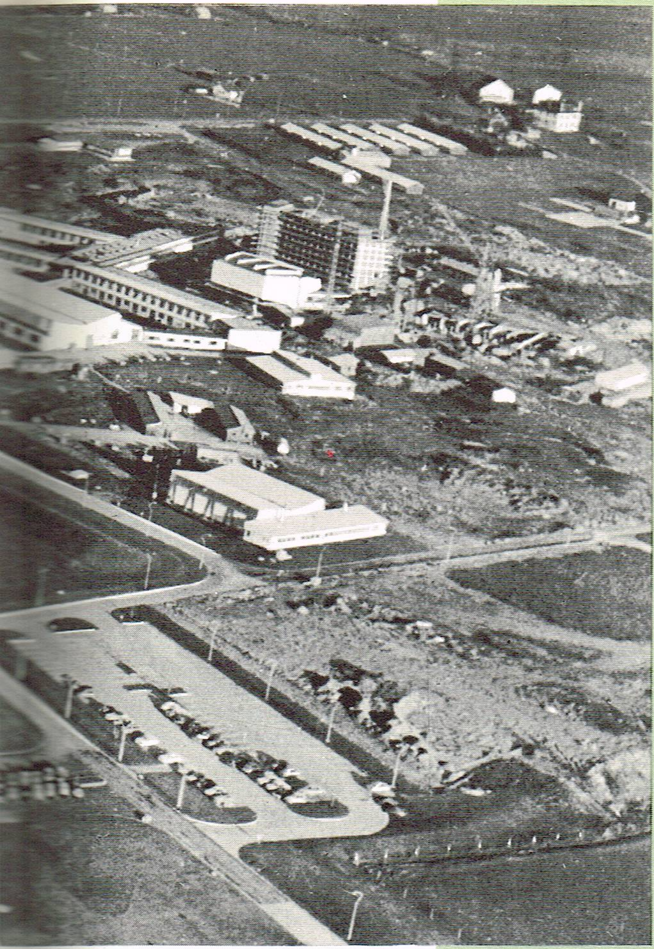
Louis LEPRINCE-RINGUET.



Vue aérienne du CERN en atterrissant à Genève. On distingue le grand cercle du synchrotron.

Maquette du grand radiotélescope de Nançay dans son état définitif.

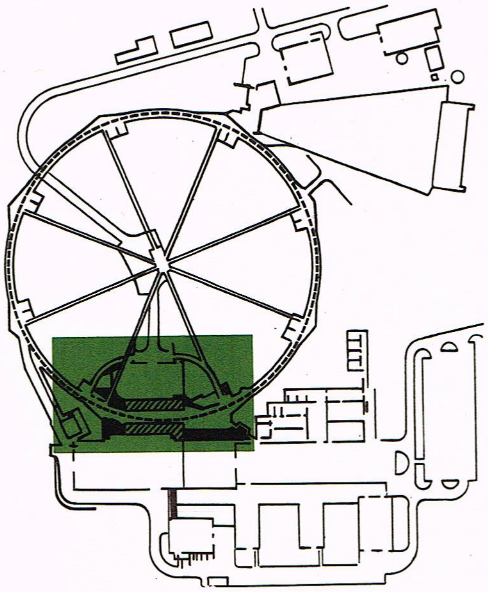




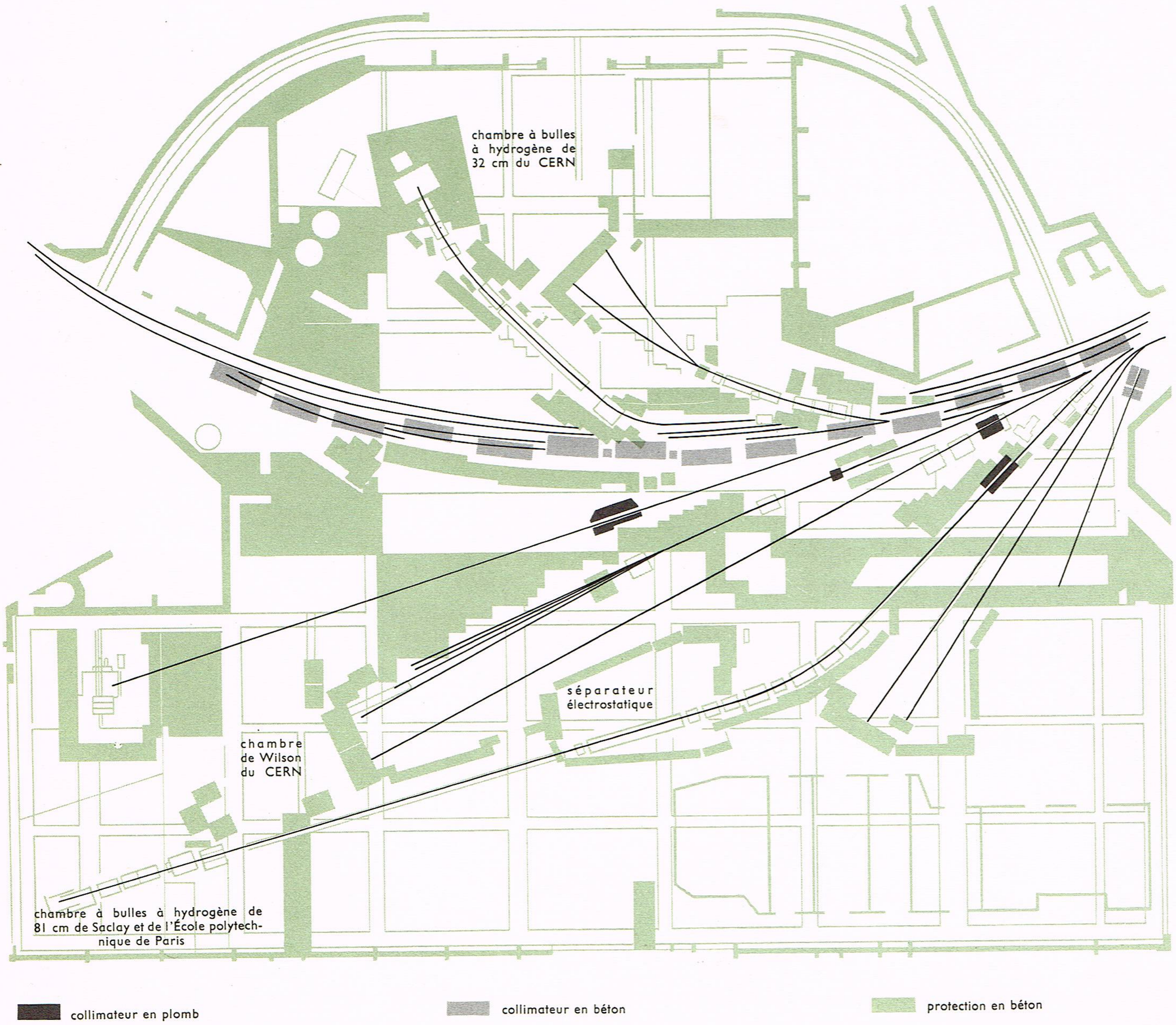
Phot. CERN, Observ. de Paris.



deux HAUTS LIEUX de la science contemporaine



Cette « gare de triage » n'est pas destinée à aiguiller des trains. Il s'agit des installations où l'on tire profit des particules accélérées par le grand synchrotron, dont une courte section est figurée ici à l'échelle. Des faisceaux secondaires dûment focalisés et déviés, sont dirigés vers les appareils de détection où seront faites les expériences. *Doc. CERN.*



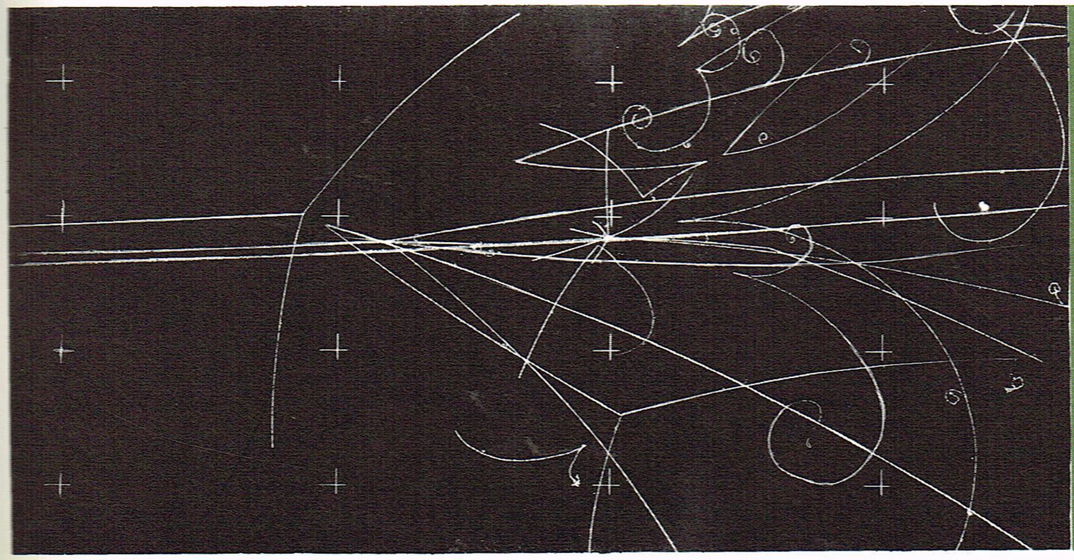
le Centre Européen de Recherche Nucléaire

Lorsqu'on atterrit sur l'aérodrome de Genève, comme cela nous arrive très fréquemment en venant de Paris par la « Caravelle » du matin, on est frappé par une sorte d'immense roue de bicyclette, que l'on aperçoit sur le terrain à proximité de l'aéroport, tout près du petit faubourg de Meyrin. Cette roue appartient à une agglomération curieuse, formée de bâtiments très modernes, dont certains présentent un aspect administratif, et d'autres une apparence plus technique. Il s'agit du Centre européen de recherche nucléaire, qui a été construit sur cet emplacement, tout contre la frontière française, entre 1954 et 1961, et qui est un des hauts lieux de la science nucléaire dans le monde. La roue que l'on aperçoit, parfaitement circulaire, est en réalité un cercle géant de 800 m de circonférence; il s'agit d'un très grand tube placé horizontalement et recouvert de béton et de terre. On voit même des rayons, qui sont effectivement des galeries se dirigeant du centre vers certains points du tube.

L'appareil qui exige un tel développement est le « synchrotron à protons ». C'est dans cette enceinte circulaire que des particules atomiques sont accélérées, acquièrent une vitesse très élevée, et finissent par frapper des noyaux d'atome, produisant ainsi des interactions nucléaires de grande énergie. Un tel appareil est extrêmement onéreux et n'est pas à la portée de n'importe quelle bourse gouvernementale; il n'en existe que de rares exemplaires dans le monde; on a dû grouper tous les pays de l'Europe occidentale pour réussir cette réalisation, car ni la France, ni l'Allemagne, ni l'Angleterre séparément ne possèdent un accélérateur aussi puissant. Les seuls que l'on puisse trouver dans le monde sont, en dehors de celui que nous allons décrire ici, un « bévatron » en Californie, à Berkeley, sur les collines qui dominent la baie de San Francisco, un « AGS » (Alternating Gradients Synchrotron) près de New York, dans l'île plate et humide de Long Island, et enfin un « synchrophasotron » à Doubna, située à 50 km au-delà de Moscou. Ainsi, chacun de ces immenses appareils alimente une partie du monde, c'est-à-dire travaille pour les physiciens qui appartiennent non seulement à un pays défini, mais à toute une région du monde : Doubna travaille pour les physiciens qui se trouvent derrière le rideau de fer, et réunit Polonais, Tchèques, Russes, Chinois, Nord Vietnamiens, etc.; Berkeley alimente les universités et les centres de recherche de la côte ouest des Etats-Unis, et Brookhaven, les nombreuses et puissantes universités de l'Est américain. Pour le Centre européen de recherche nucléaire, ce sont toutes les universités d'Europe qui se trouvent intéressées au fonctionnement de cette machine et qui vont profiter des irradiations que l'on va effectuer à Genève.

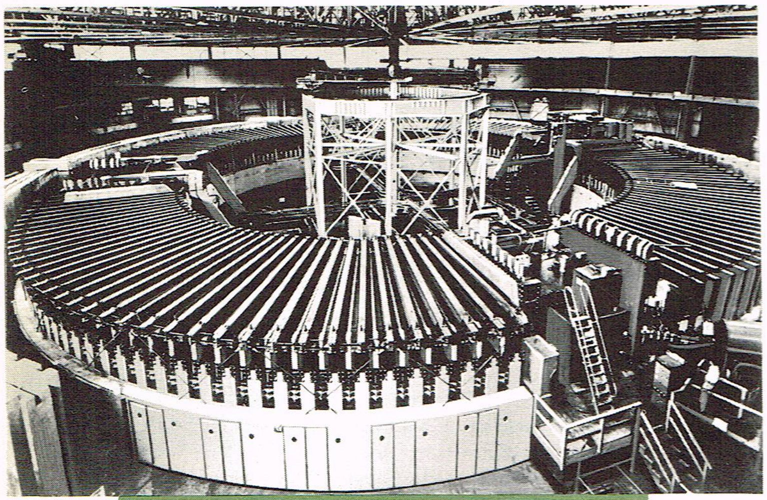
Proposons-nous de visiter le CERN. L'entrée est facile; ce n'est pas un centre atomique comme Los Alamos ou Livermore, dans lesquels on ne pénètre que difficilement, avec des autorisations policières. Ici, rien n'est caché : c'est de la science de base, et l'un des principes du CERN est que tout est publiable ouvertement; il n'y a pas de problème secret.

Donc, nous entrons par le poste principal et nous nous trouvons devant une série de grands bâtiments, dont le plus curieux est à coup sûr celui du synchrotron, qui est extrêmement étendu. Nous y arriverons tout à l'heure, mais, auparavant, nous allons rendre visite à un accélérateur de particules moins gigantesque,

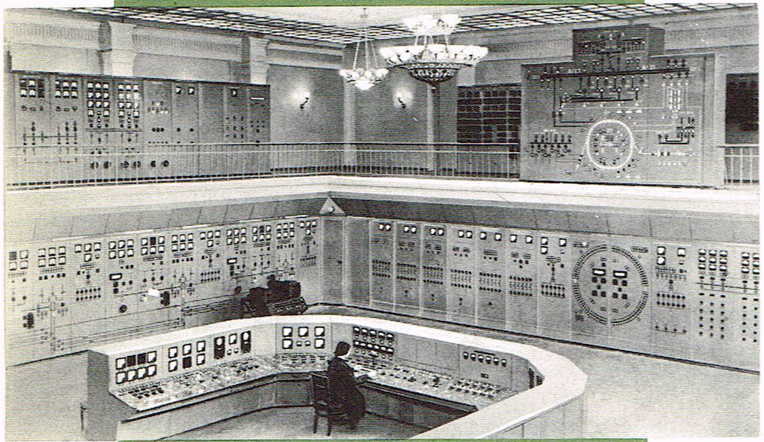


Interaction de grande énergie dans la chambre à bulles B P 3 de l'Ecole polytechnique. Les quatre particules rapides dont les traces, venant de la gauche, sont parallèles sont des mésons π de 18 milliards d'électron-volts, provenant du grand synchrotron du CERN. On voit l'un de ces mésons produire une interaction sur un noyau du liquide de la chambre; deux traces épaisses, donc lentes, partent de cette interaction : un proton et un méson π négatif. Mais beaucoup d'autres phénomènes apparaissent plus loin : ils proviennent de particules électriquement neutres (mésons π^0, mésons K^0, neutrons) dont la trace n'est pas visible; ces particules neutres se manifestent lors de leur désintégration en particules chargées ou encore lors de leur interaction avec un autre noyau.

Phot. Ecole polytechnique.



Le bévatron de Berkeley (Etats-Unis).



Le synchro-phasotron de Doubna (U. R. S. S.).

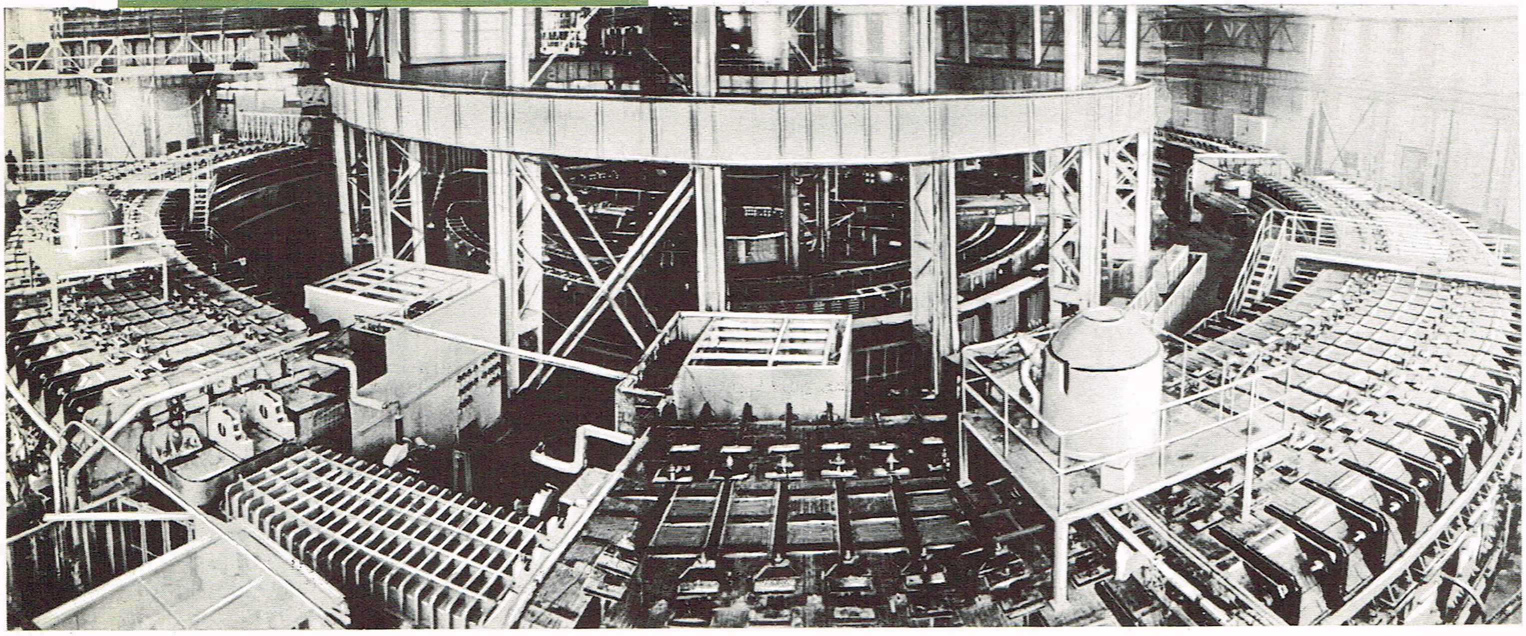
***Ci-dessus*, la salle des commandes.**

construit au CERN quelques années avant le synchrotron, et qui permet de faire de la très bonne physique, sans atteindre toutefois les énergies du grand appareil. Il s'agit d'un cyclotron cette fois, et même d'un synchro-cyclotron. On l'appelle le « S. C. ». Il n'y a d'ailleurs que deux accélérateurs de particules au CERN : le « S. C. » et le « P. S. » (synchrotron à protons.)

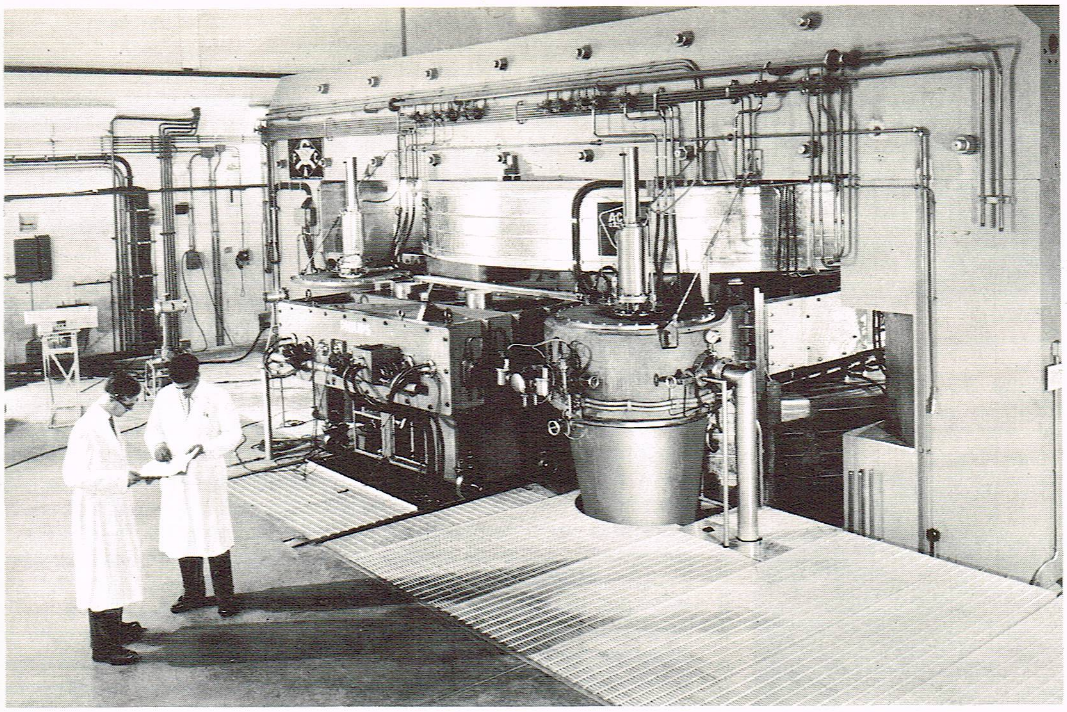
le synchro-cyclotron

Cet appareil est le dernier d'une série de synchro-cyclotrons dont les constructions se sont échelonnées entre 1945 et 1960. Ce sont des appareils qui furent tout à fait à la pointe de la physique des particules fondamentales après la guerre. Le premier fut construit à Berkeley, en Californie, dans le Radiation Laboratory; il a même servi, avant d'être utilisé pour la physique des particules élémentaires, à séparer les isotopes de l'uranium vers la fin de la guerre, lorsque les Américains ont voulu fabriquer un petit peu d'uranium 235 très pur. Tous les progrès dans la physique des mésons π et des interactions nucléaires ont été faits grâce à ce genre d'appareil, jusqu'en 1958 environ. Les synchrotrons à protons, version plus moderne et de conception différente, sont arrivés au cours des dernières années seulement.

Le synchro-cyclotron donne une impression de puissance et de concentration; on y voit d'énormes cylindres de fer, qui sont les pôles d'un immense électro-aimant, mais tout cela n'est pas très étendu dans l'espace. Nous sommes habitués aux électro-aimants de petites et moyennes dimensions, mais ici les pièces polaires ont une dizaine de mètres de diamètre, ce qui est énorme. Songez combien il est difficile d'avoir, dans l'industrie, des machines-outils capables de fabriquer des cylindres de ce diamètre. On ne peut pas aller plus loin, on est limité par les possibilités industrielles et on utilise là les plus gros électro-aimants du monde. Tout se passe à l'intérieur : les deux pièces polaires sont des cylindres assez plats, dont l'axe est vertical, si bien que le champ magnétique est vertical lui aussi. Ce champ est très intense; il atteint entre 15 000 et 20 000 gauss; l'entrefer de cet aimant est d'un mètre environ, et c'est dans le volume de l'entrefer que tout va se passer. On est très loin du cercle de 800 m de tour.

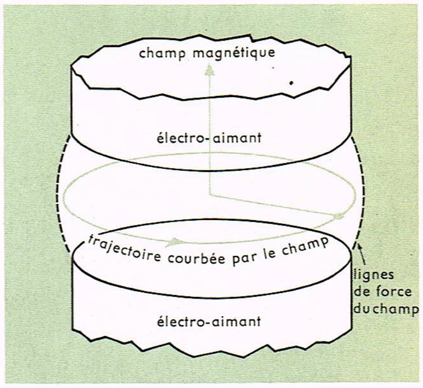


Phot. U. S. I. S., B. S. I.



Le synchro-cyclotron du CERN : 650 millions d'électron-volts.

Ci-dessous, **accélération des particules dans un cyclotron.**



Dès l'abord, on doit se familiariser avec les problèmes de champ magnétique. Nous connaissons celui de la Terre, qui détermine l'orientation des boussoles; nous connaissons également les champs magnétiques plus intenses qui permettent aux machines électriques classiques de fonctionner. Ces notions ne sont pas récentes; elles ont été développées au début du XIXe siècle, et la grande science de l'électromagnétisme a été mise au point entre 1820 et 1870. Il faut dire que les accélérateurs de particules font un très large appel à cette science et lui donnent un caractère moderne. Nous aurons donc à nous préoccuper de revoir les notions de base de l'électromagnétisme, de bien les comprendre, afin de suivre les développements de la physique moderne; les explications correspondantes seront fournies dans la suite de cet ouvrage.

C'est dans ce champ magnétique uniforme, dont les lignes de force sont verticales, que l'on va introduire, par la région centrale, des particules atomiques. Là encore, on peut se demander comment on introduit ces particules et de quelle nature elles sont. Il s'agit de protons. Les protons sont les plus petits noyaux atomiques appartenant aux atomes les plus légers, à savoir ceux de l'hydrogène. Il est maintenant bien connu que les atomes sont constitués d'un noyau central et d'un nuage d'électrons périphériques, le noyau central étant, en général, caché au regard des atomes voisins : quand des atomes se rencontrent, ils ne voient que leurs nuages périphériques, les noyaux étant bien masqués et ne pouvant être perçus par les autres atomes. C'est pour cela que les interactions « chimiques » ou physiques habituelles, telles que les réactions d'oxydation et de combustion, s'effectuent sans que les noyaux d'atome aient à intervenir; c'est pour cela qu'après une oxydation ou une combustion les éléments chimiques se retrouvent toujours dans les mêmes proportions. Si le fer est à l'état pur ou à l'état rouillé, c'est toujours du fer, et l'on peut récupérer, en traitant convenablement la rouille par des moyens chimiques, tout le fer qui a subi l'oxydation.

Ainsi nous disposons d'hydrogène et non de protons, et pour obtenir ces derniers il faut agir d'une certaine façon sur l'hydrogène pour arracher son nuage d'électrons. Dans le cas simple de l'hydrogène, il n'y a qu'un électron, si bien que si l'on arrive à le séparer par une sorte d'arrachage violent, le proton (c'est-à-dire le noyau épluché, comme on dit encore) restera tout seul et, puisqu'il est électrisé (car tout noyau est chargé d'électricité positive), on pourra l'accélérer avec des champs électriques.

Nous voyons donc apparaître des notions nouvelles : arracher des électrons d'un atome s'appelle « ioniser » l'atome et, en physique atomique ou nucléaire, l'ionisation est une activité courante. Nous pouvons déjà annoncer qu'avec les rayons ultraviolets, ou les rayons X, ou la radio-activité, on provoque facilement l'ionisation, mais on peut aussi la provoquer par des étincelles ou des effluves, et c'est ainsi que l'on crée l'ionisation nécessaire à l'obtention des protons dans les accélérateurs. Donc, pour comprendre ce qui se passe dans le synchrocyclotron, nous avons fait appel à quelques notions, à savoir : champ magnétique, constitution de l'atome, et enfin obtention du noyau dépouillé de ses électrons par ionisation.

Ainsi nous pouvons dire que nous avons au centre d'un immense électro-aimant des protons dans un petit espace. On établit tout d'abord un champ électrique, horizontal cette fois, pour commencer à accélérer ces protons. Nous pouvons bien comprendre que nous avons ainsi une bouffée de protons qui vont tous se propager ensemble dans la même direction sous l'action de ce champ électrique, car on sait qu'un champ électrique agit sur des particules électriques, soit en les attirant, soit en les repoussant. Chaque proton va prendre ainsi une certaine énergie, qui ne sera pas bien grande. Nous désignerons

Phot. CERN.

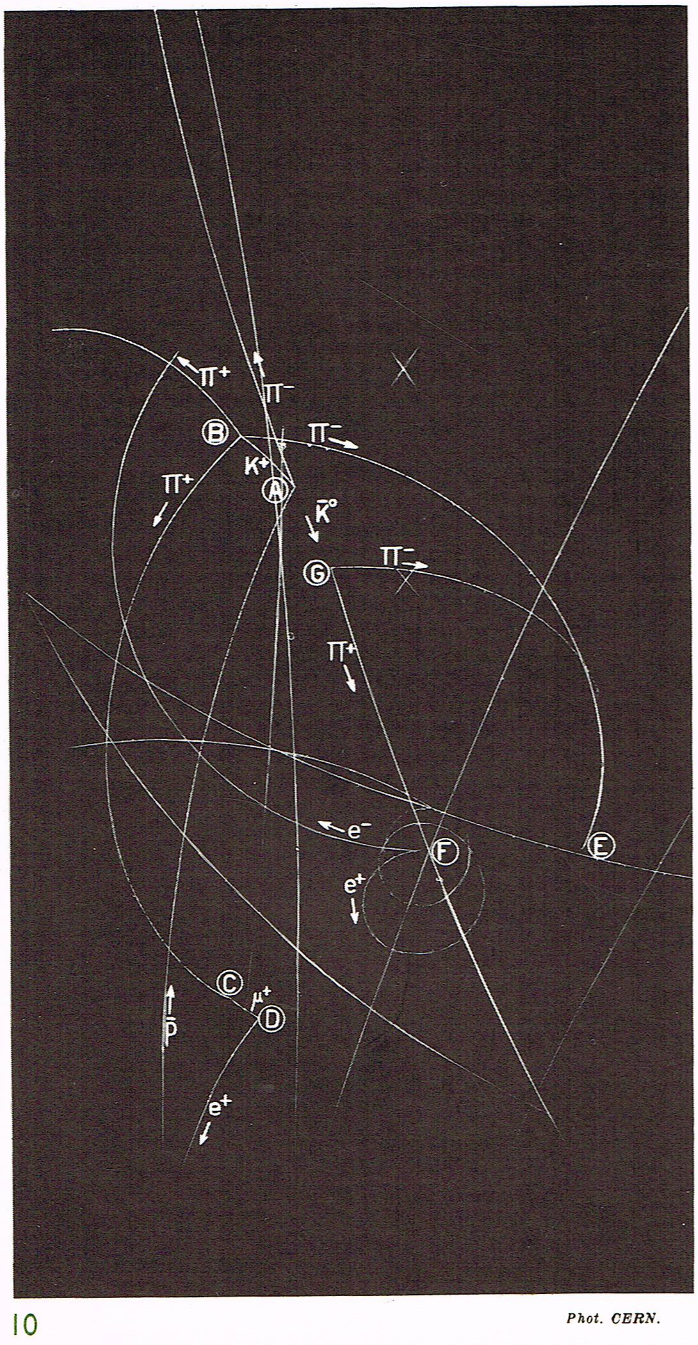
1 électron-volt = énergie prise par un électron ou un proton soumis dans le vide à une différence de potentiel de 1 volt.

Synchro-cyclotron : 650 millions d'électron-volts.

Cosmotron de Berkeley : 6 milliards d'électron-volts.

Synchrotron du CERN et AGS de Brookhaven : 28 milliards d'électron-volts (record actuel).

Un antiproton lent (trace $\overline{p}$) venant du bas sur la figure s'arrête en A.
Il s'annihile alors avec un proton de la chambre à bulles à hydrogène. Cette annihilation provoque la création de mésons π et de mésons K. On observe les désintégrations spontanées de ces derniers ($\overline{K^0}$ et K^+) en mésons π. Les traces sont courbées par le champ magnétique en raison inverse de leur « quantité de mouvement ».



Phot. CERN.

les énergies avec une unité bien connue par les physiciens, qui s'appelle un « électron-volt », et nous dirons que chaque proton de la bouffée va prendre une énergie de quelques dizaines ou de quelques centaines d'électron-volts grâce à un champ électrique établi au bon moment. Mais nous voulons donner au proton une énergie beaucoup plus grande : 600 millions d'électron-volts, et nous ne sommes pas au bout de nos peines. Comment parvenir à cette énorme énergie, et quelles vont être les notions physiques qu'il va falloir faire intervenir pour comprendre le fonctionnement d'un tel appareil? Le plus simple est peut-être d'abord de fermer les yeux sur tout ce qui est intermédiaire et de voir le phénomène final : une seconde après le démarrage du faisceau de protons, nous ouvrons à nouveau les yeux et nous voyons qu'une bonne partie de ces protons sont maintenant au bord extérieur du gros aimant, circulant à toute allure avec une vitesse presque égale à celle de la lumière, soit 300 000 km par seconde, toujours groupés en peloton et s'apprêtant à être largués vers l'extérieur comme les projectiles d'une immense fronde, pour venir heurter avec une grande violence une cible, par exemple un morceau de cuivre ou de tout autre matériau, dont des noyaux vont être désintégrés sous l'action du faisceau énergique de protons. Tout n'est pas facile à comprendre entre le début et la fin de l'opération. L'ensemble des phénomènes qui se passent à l'intérieur du gros aimant se développe dans le vide, car il ne faut pas que les protons puissent rencontrer beaucoup d'atomes sur leur chemin, sinon leur direction deviendrait moins bien définie et l'on ne pourrait pas les accélérer convenablement sous forme de faisceau. D'autre part, une fois mis en mouvement par le premier champ électrique, ils ont tendance à se diriger en ligne droite. Comme on veut les accélérer à l'intérieur de l'aimant par un dispositif circulaire, il faut obliger ces protons à ne pas s'échapper; il faut donc incurver leur trajectoire. C'est pour cela d'ailleurs que l'aimant est fait et qu'il possède un champ magnétique vertical uniforme. Les trajectoires des particules électrisées sont analogues à de petits éléments de courant électrique, et nous avons tous appris en classe qu'un champ magnétique agit sur un courant : c'est le grand savant français Ampère qui, autour de 1820, a énoncé le premier les lois de l'action des champs magnétiques sur les courants. Ainsi, les particules en mouvement sont comme de petits éléments de courant, et leurs trajectoires subissent l'action d'un champ magnétique avec la même règle, celle du tire-bouchon ou celle dite « des trois doigts », bien connue des manuels scolaires. Voici ci-contre un exemple de cette action.

S'il n'y avait que le champ magnétique et le faisceau de protons mis en mouvement comme nous l'avons vu, il ne se passerait rien, c'est-à-dire que le faisceau de protons resterait, avec son énergie, sur une trajectoire circulaire dont le rayon de courbure serait plus ou moins grand, selon la valeur du champ magnétique : si l'on établit un champ magnétique fort, le rayon de courbure sera tout petit; si le champ magnétique est faible, le rayon de courbure sera plus grand. Comme le champ magnétique est fort, le faisceau sera très courbé au départ. Mais nous voulons à chaque tour, ou même deux fois par tour, donner une accélération au faisceau, c'est-à-dire lui fournir une impulsion qui sera nécessairement une impulsion électrique puissante. Alors, il faut nous arranger pour que le faisceau

soit soumis, chaque demi-tour, à une force électrique qui donnerait une vitesse supplémentaire à chaque proton, et pour cela on adopte le principe du cyclotron classique, celui qui a été inventé par Lawrence en 1930. L'appareil qui se trouve dans le vide et qui remplit tout l'entrefer de l'électro-aimant se compose en réalité de deux morceaux appelés des « D ». En effet chaque morceau ressemble à un « D » majuscule plus ou moins régulier. On peut se représenter une immense boîte à camembert que l'on aurait coupée verticalement selon un de ses diamètres, après quoi on aurait écarté un peu les deux parties. Ces deux parties sont en réalité métalliques, et c'est entre elles que l'on établit au moment convenable un champ électrique intense et très bref, à savoir un champ de haute fréquence. On comprend donc ce qui va se passer; la figure le représente bien. Le faisceau de protons, en arrivant dans la partie qui est située entre les deux « D », va ressentir les effets d'un champ électrique intense et bref dirigé de façon à l'accélérer, après quoi, pénétrant à l'intérieur de la boîte métallique, les particules ne seront plus sensibles à aucun champ électrique : dans une cage de Faraday, en effet, on le sait bien, les champs électriques extérieurs n'agissent pas. En revanche, le champ magnétique agira toujours; ce n'est pas un petit morceau de cuivre qui l'empêche d'exercer son action. Donc, les particules vont tourner, et leur rayon de courbure sera de plus en plus grand à mesure qu'elles seront accélérées; lorsqu'elles vont se présenter de l'autre côté entre les deux « D », on aura eu le temps de changer le sens du champ électrique si la fréquence de ce champ est bien accordée avec la vitesse des particules. Elles vont donc se sentir soumises à un nouveau champ électrique qui va les accélérer de nouveau, puis vont rentrer à l'intérieur du « D » et ne seront plus sensibles qu'au champ magnétique. Ainsi le faisceau de particules va progressivement décrire des demi-cercles dont le rayon sera de plus en plus grand. Dans le cyclotron ordinaire, le temps que mettent les particules à décrire leur demi-cercle est toujours le même : pour les petits cercles, la vitesse est moins grande et la longueur est aussi moins grande; pour les grands cercles, la vitesse est plus grande et la longueur aussi, mais le temps est toujours le même. Par conséquent, si l'on fait agir aux bons moments un champ électrique de haute fréquence qui soit bien défini, il n'y a pas besoin de changer cette fréquence au cours du temps, et les particules seront toujours accordées à ce champ, c'est-à-dire retrouveront toujours la petite pichenette qui leur donnera un peu plus d'énergie à chaque passage. Finalement, au bout de plusieurs milliers de révolutions, les particules arriveront toujours groupées à l'extérieur de l'aimant; elles seront alors accélérées au maximum.

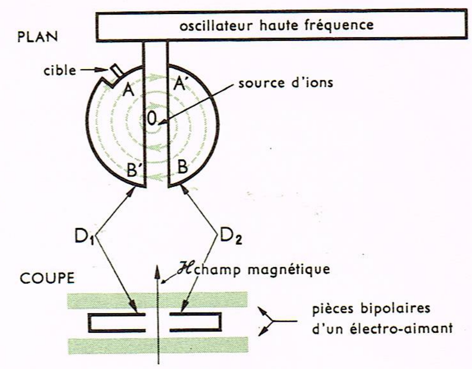


Schéma de principe du cyclotron. Les ions, injectés en O, décrivent une spirale. Ils sont accélérés à chaque passage entre les deux D par une impulsion électrique, puis sortent et frappent la cible C.

premier contact avec la relativité

Jusqu'à présent, nous n'avons pas fait appel à autre chose qu'à l'électromagnétisme classique : l'action des champs électriques ou magnétiques sur les petits éléments de courant. Mais le problème devient un peu plus difficile quand les énergies communiquées aux particules sont grandes. En effet, quand on donne de l'énergie à une bille, sa vitesse augmente; lui donner plus d'énergie correspond à augmenter davantage sa vitesse. L'énergie cinétique, ce qui signifie l'énergie due au mouvement, est proportionnelle au carré de la vitesse. Or, on s'aperçoit dans un grand accélérateur de particules que la vitesse des particules augmente au début, lorsqu'on les accélère, puis augmente de moins en moins quand on continue à leur communiquer de l'énergie; finalement, elle n'augmente plus du tout, ou d'une façon imperceptible, lorsque l'énergie devient très grande, et pourtant cette énergie augmente constamment, puisque à chaque tour on donne une énergie supplémentaire aux particules. A la fin, elles font le tour avec une vitesse constante et une énergie croissante. Il y a là une notion qui n'est plus classique et qui ne peut se comprendre que si l'on fait appel à la relativité introduite par Einstein autour de 1910.

Le noyau atomique est fait de protons et de neutrons, mais les physiciens ont découvert à ce jour l'existence d'une trentaine de particules élémentaires que l'on peut classer en cinq groupes :

- **photon** — C'est l'unité ou quantum de rayonnement électromagnétique (lumière, rayons X et gamma);
- **leptons (1)** — Électron, positon, neutrino, antineutrino, muon;
- **mésons (2)** — Mésons K, mésons π;
- **nucléons (3)** — Proton, neutron, antiproton, antineutron;
- **hypérons (3)** — Particules lambda et antilambda, xi et antixi, sigma et antisigma.

(1) particules légères;
(2) particules de masse moyenne;
(3) particules les plus lourdes, raison pour laquelle on les appelle aussi « baryons ».

La mécanique classique n'explique pas cela, mais la mécanique relativiste prévoit qu'il y a une vitesse limite pour toute particule matérielle, que cette vitesse est égale à la vitesse de la lumière dans le vide, c'est-à-dire 300 000 km par seconde, et que la masse de la particule augmente avec l'énergie. Et encore nous ne considérons que le synchro-cyclotron; mais lorsque l'énergie des protons atteindra 25 milliards d'électron-volts au lieu des 600 millions du synchro-cyclotron, le domaine sera encore plus profondément relativiste, et au cours de presque toute l'accélération les protons conserveront pratiquement la même vitesse.

Voilà donc un premier contact, bien curieux, avec la relativité; nous aurons affaire à cette théorie constamment, et il faut dire dès l'abord que la relativité d'Einstein est pour le physicien un guide extrêmement sûr; nous retrouvons les résultats de la relativité dans toutes les réactions nucléaires. Les bilans d'énergie nucléaire ne sont vrais que si l'on utilise les formules relativistes, en particulier l'équivalence de la masse et de l'énergie. C'est pour nous comme une sorte de balance de précision d'une sécurité absolue. Nous allons retrouver cela bientôt, lorsque nous examinerons ce que l'on fait avec les accélérateurs de particules.

protons, neutrons, mésons, photons...

Voici donc le schéma de principe du synchro-cyclotron : les particules accélérées vont ainsi frapper une cible, et comme ce sont des noyaux qui ont été mis en mouvement, ce sont des noyaux également qu'ils vont heurter. On peut même dire que, dans ces noyaux, ce sont des protons et des neutrons qui seront les victimes des chocs des protons incidents : nous allons donc pouvoir étudier les interactions entre protons et protons, ou protons et neutrons. Nous aurons à nous initier à la constitution des noyaux d'atome, pour bien comprendre comment les protons et les neutrons peuvent constituer les noyaux stables de notre matière, comment ils sont liés ensemble, comment on peut les extraire du noyau. Même lorsque nous connaîtrons cela, nous ne serons qu'au début de la connaissance nucléaire. Les particules fondamentales vont beaucoup plus loin, et nous verrons que l'on ne provoque pas seulement des collisions qui amènent l'explosion des produits existant dans le noyau; on va beaucoup plus avant, car, avec l'énergie disponible dans ces collisions, on parvient à créer des particules nouvelles qui ne préexistaient pas. Grâce au synchro-cyclotron que nous venons de décrire, on accélère des protons qui vont pouvoir, par collision avec d'autres noyaux, non seulement les faire exploser, mais créer des particules nouvelles : les « mésons ». Il y en a de plusieurs espèces, mais seuls les mésons π vont être créés avec l'accélérateur que nous avons décrit, les autres — à savoir les mésons lourds — seront, au contraire, créés avec le synchrotron à protons que nous irons voir tout à l'heure.

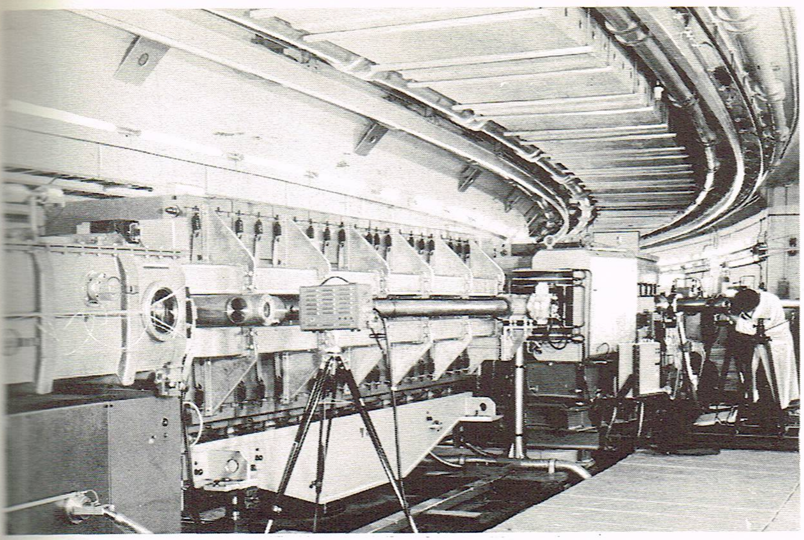
Nous voilà donc devant un nouveau problème : qu'est-ce qu'un méson π? comment un méson π est-il connecté aux interactions entre protons et neutrons? quelle est la signification du méson π, particule tout à fait fondamentale, mais instable? comment l'a-t-on découvert? comment peut-on le créer et l'observer?

Nous verrons, et c'est là également une très fructueuse comparaison, que le méson π correspond au champ nucléaire : c'est une particule dont on peut dire qu'elle constitue l'aspect corpusculaire du champ nucléaire. Cette phrase est peut-être un peu mystérieuse, mais on peut la rapprocher d'une autre phrase plus intuitive qui est la suivante : le photon, c'est-à-dire le grain de lumière, représente l'aspect corpusculaire du champ électromagnétique. Le photon, c'est la lumière; nous verrons, en effet, que la lumière se manifeste sous un aspect granulaire; le phénomène photo-électrique en est la preuve la plus évidente. Le grain de lumière ne préexiste pas dans l'univers, mais la lumière se produit lorsque les interactions électriques ou magnétiques ont lieu; dans un tube au néon, il y a de la lumière parce que les électrons bougent et interagissent les uns avec les autres : c'est le champ électromagnétique. La lumière ne préexiste pas dans un tube de néon; elle se produit lorsqu'on met le courant, c'est-à-dire lorsqu'on fait bouger les électrons, et ce sont des modifications d'énergie des électrons qui provoquent l'émission des photons. Pour les noyaux, il ne s'agit plus du champ électromagnétique : les protons et les neutrons sont très fortement liés dans un noyau par autre chose qu'une attraction électrique ou magnétique : **c'est une attraction nucléaire,** et l'on dit qu'ils baignent dans le champ nucléaire de leurs voisins. Ce champ nucléaire se traduit par un corpuscule qui est le méson π, imaginé par un physicien japonais, Ideki Yukawa, en 1936.

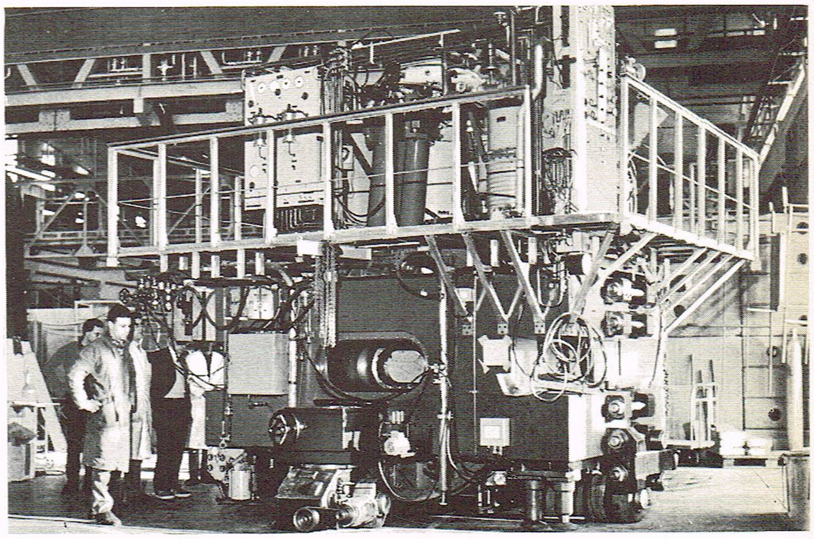
Que le lecteur ne se rebute pas devant ces notions difficiles; comprendre la physique, même dans ce qu'elle a de très élémentaire, est passionnant, mais difficile. Il va falloir se familiariser avec les notions de champ et de particule qui vont se trouver expliquées tout au long de cet ouvrage; tout d'abord les champs classiques, puis la relativité, avec ses lois extraordinairement simples et claires, mais stupéfiantes, puis enfin la mécanique ondulatoire ou quantique, qui va tenir compte du fait que la lumière est constituée par des corpuscules et par des ondes, ce qui explique à la fois l'existence des photons aussi bien que celle des interférences, qui va tenir compte également du fait que les électrons sont des petits projectiles, des petits grains, mais sont aussi, de façon contradictoire, des ondes, et qui va promouvoir tout un formalisme nouveau pour expliquer la réalité telle qu'elle se présente à nous et qu'elle va être décrite dans les pages qui suivent.

La grande salle dans laquelle se trouve le synchro-cyclotron est remplie de matériel; on y voit beaucoup d'électronique avec une quantité impressionnante de « racks » (armoires aux tiroirs remplis de circuits électroniques et montés sur roulettes), de tubes qui scintillent plus ou moins vite, de dispositifs compliqués qui surprennent sans qu'on puisse bien comprendre leur utilité. On voit également de grands appareils protégés par d'imposantes structures de béton ou de métal. Tout cela est fait pour la détection des particules qui sont produites par les protons rapides lorsqu'ils interagissent avec les noyaux d'une cible. Mais nous allons retrouver ces appareils beaucoup plus nombreux et beaucoup plus énormes dans les grands halls qui entourent le synchrotron à protons. Donc, nous dirons seulement qu'il s'agit d'ensembles permettant la détection des phénomènes soit avec des compteurs d'électrons, soit avec des scintillateurs, soit avec des compteurs Cerenkov, soit avec d'autres dispositifs électroniques, ou bien qu'il s'agit d'autres sortes d'appareillages, à savoir les grandes chambres à bulles, admirables outils pour l'obtention de photographies des phénomènes, et dont l'emploi est courant avec les grands accélérateurs de particules.

Dirigeons-nous maintenant vers l'anneau du synchrotron à protons.



Le hall du synchrotron où les faisceaux secondaires de particules sont séparés.



La chambre à bulles à hydrogène liquide fontionne à la température de — 248 °C.

le synchrotron de 28 milliards d'électron-volts

L'anneau du synchrotron est entouré de trois ou quatre immenses halls remplis de matériel d'électronique, de pièces de mécanique, de générateurs divers, de blocs de béton, de coffres métalliques et d'un nombre impressionnant d'électro-aimants de dimensions moyennes et de forme curieuse, appelés soit « aimants de déviation », soit « quadrupoles ». On observe dans ces grands halls, qui ont plus de 150 m de longueur, des alignements de toute une série d'aimants, puis de tubes, puis d'autres aimants, puis de tubes encore dans une direction un peu différente, puis de grands blocs de béton, puis une énorme masse de fer ou de plomb avec une toute petite fente bien alignée sur l'axe des tubes précédents : tout cela représente certainement un travail de préparation pour des faisceaux de particules qui vont sortir de l'accélérateur. On a immédiatement l'impression, en traversant ces halls avant d'arriver au sanctuaire du synchrotron, que l'accélérateur va fournir des faisceaux secondaires que l'on devra séparer, définir selon certains critères et envoyer très loin, bien en dehors des contaminations du reste, dans un appareil très protégé. C'est bien cela : l'un de ces dispositifs va séparer, parmi l'ensemble des produits secondaires, les rares antiprotons rapides que vont fournir les interactions nucléaires, et ces antiprotons rapides, précieusement sélectionnés dans ces dispositifs de faisceaux, vont être envoyés dans un appareil qui permettra d'étudier leur comportement lorsqu'ils rencontreront des noyaux d'atome.

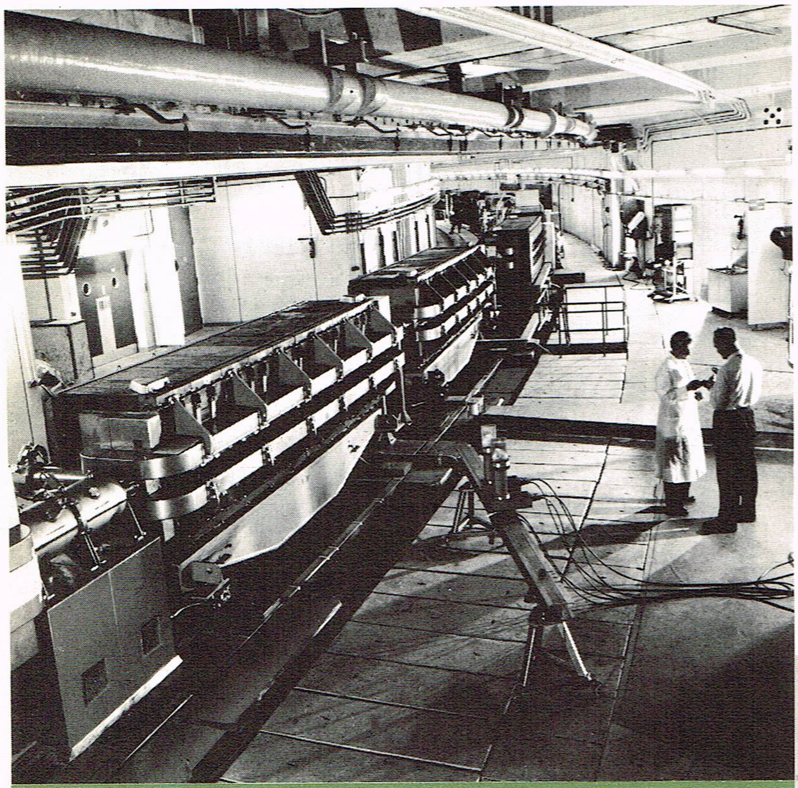
L'un de ces appareils, le plus utilisé, sera une grande chambre à bulles à hydrogène liquide dans laquelle entreront des antiprotons rapides : ces antiprotons interagiront avec les noyaux d'hydrogène du liquide de la chambre et donneront des phénomènes de la plus haute importance pour la compréhension de la physique des particules et des antiparticules. Nous sommes frappés par l'extrême précision de la mise au point de ces longs faisceaux : celui des antiprotons rapides a près de 150 m de long, et il se termine par une petite fente de 6 mm de haut seulement. Cela signifie que ces antiprotons sont parfaitement guidés sur toute cette longueur et que, dans cette petite fente de 6 mm sur 1 m de profondeur, ils doivent pénétrer eux seuls, les mésons K ou les mésons π étant éliminés, probablement difficilement, du faisceau. C'est bien exact. Les énormes énergies de plusieurs milliards d'électron-volts communiquées aux particules même secondaires sont telles, que l'on a bien du mal à séparer les antiprotons des mésons π ou des mésons K négatifs qui les accompagnent en grand nombre. Plus l'énergie est grande, plus le champ magnétique à utiliser pour la séparation doit être important et agir sur une longueur plus considérable; au bout de 150 m, on parvient à réaliser une séparation de quelques millimètres seulement entre les antiprotons rapides et les mésons K rapides. On devine que la mise au point d'un faisceau d'antiprotons exigera, en plus des moyens mécaniques et des appareillages coûteux, beaucoup de temps de la part d'un grand nombre de physiciens. Il a fallu un an avec une bonne demi-douzaine de physiciens pour préparer un tel faisceau.

Nous reviendrons plus tard aux chambres à bulles et aux scintillateurs. Franchissons les murs épais de béton et présentons-nous à l'entrée de l'accélérateur lui-même. Que voyons-nous? Un grand tunnel circulaire forme l'entourage : on a commencé par faire ce tunnel, et même on a pensé que si l'accélérateur ne donnait pas satisfaction on transformerait ce tunnel circulaire en un vélodrome souterrain ou encore en une piscine où toutes les distances habituelles pourraient être nagées sans virages. Ce tunnel est assez grand, un peu plus petit néanmoins que celui du métro; on peut en faire le tour très facilement et l'on aperçoit çà et là les petits tunnels rectilignes qui vont vers le centre. Lorsqu'on a construit le synchrotron, une fois le tunnel aménagé, on a dû réaliser une base à la fois solide et précise, un socle de béton sur lequel devait être placé le synchrotron.

Mais qu'est-ce qu'un synchrotron et quelle serait la différence entre synchrotron et cyclotron?

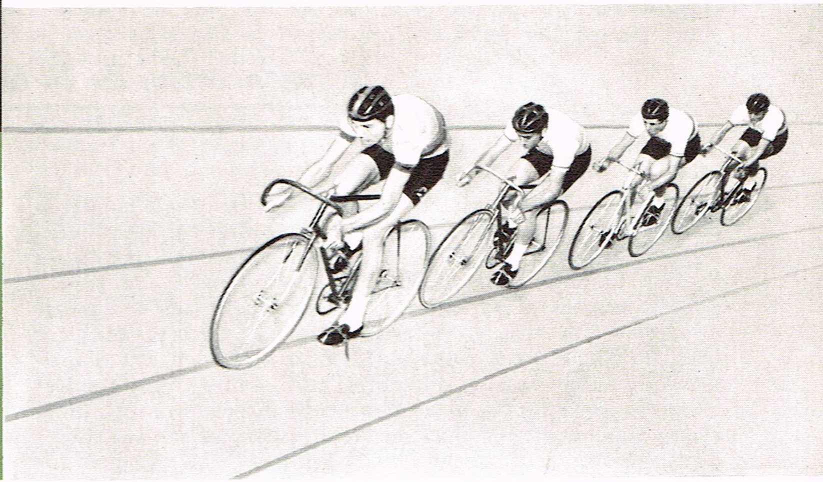
Le principe est tout différent, et, pour mieux le faire comprendre, nous allons faire appel à une image familière, celle des pelotons

Phot. CERN.



Le tunnel du synchrotron.

tours suivants. Il ne faut pas non plus que ceux qui ont pris un peu d'avance se voient accélérés davantage, car le peloton se disloquerait, et l'on ne pourrait plus alors agir sur lui par des impulsions commandées au bon moment. Ce schéma donne le principe du synchrotron : on accélère un faisceau de particules tout d'abord sous forme rectiligne, à l'aide d'un accélérateur linéaire, et, pour ce faire, dans un petit récipient, on produit des protons par ionisation de l'hydrogène, comme pour le synchro-cyclotron, puis on accélère ces protons avec un champ électrique. Ce premier champ leur communique une petite vitesse, mais on reprend cette accélération dans une série de dispositifs successifs, si bien que le peloton de particules se trouve soumis à toute une série de poussées successives dans le même sens et prend ainsi une grande vitesse : à la fin de l'accélération linéaire il a déjà acquis 40 millions d'électron-volts, ce qui est beaucoup, mais encore très loin des 25 milliards d'électron-volts que l'on veut communiquer à ce peloton. A cette énergie, la vitesse des protons est déjà notable, de l'ordre de 100 000 km à la seconde, mais elle est encore loin de la vitesse maximale que des particules peuvent prendre, c'est-à-dire de la vitesse de la lumière que nous avons déjà rencontrée précédemment.

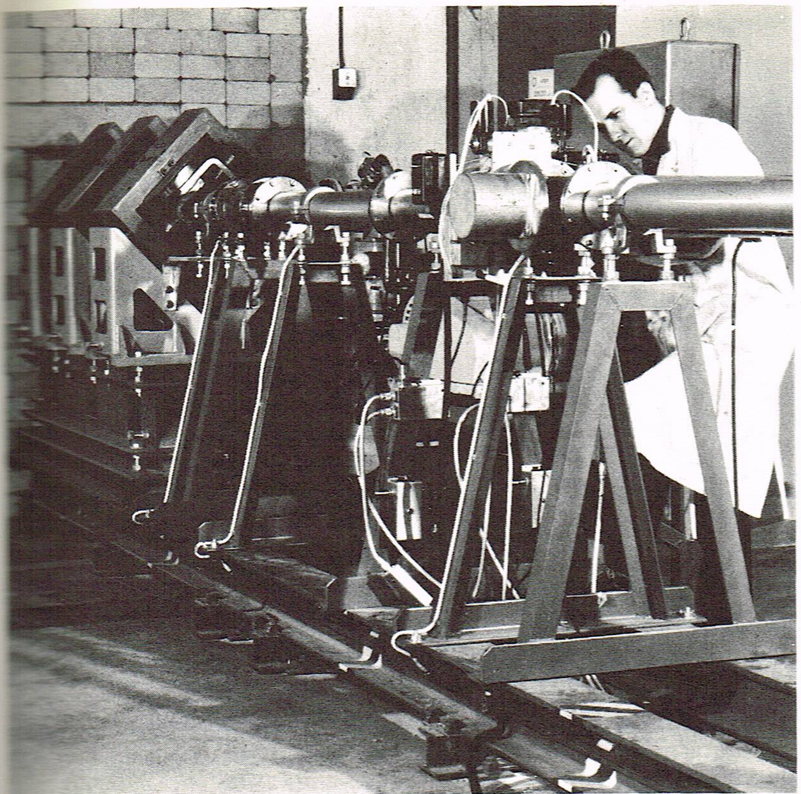


Coureurs cyclistes photographiés dans le virage d'un vélodrome. Dans le synchrotron, le peloton de quelques coureurs cyclistes groupés est remplacé par un peloton de 400 milliards de protons groupés. Au lieu de décrire en une heure les quelques dizaines de tours de vélodrome que décrivent les cyclistes, le peloton de protons groupés décrit en moins de 2 secondes plus de 300 000 km.

de coureurs cyclistes. Imaginons un groupe de coureurs cyclistes installés au départ immobiles sur leurs bicyclettes. On va se proposer de les faire aller de plus en plus vite en leur communiquant de temps en temps une petite poussée; on commencera par les accélérer en ligne droite, c'est-à-dire qu'une première partie de piste sera rectiligne. Ils prendront le départ, et seront poussés progressivement par des impulsions successives : on peut imaginer que ce seront des courants d'air puissants successifs qui augmenteront la vitesse de tous les cyclistes du peloton simultanément. Dans cette première phase, les cyclistes seront accélérés linéairement : tous les 10 m, par exemple, un souffle violent dans le dos leur donnera un peu plus de vitesse. Après cette première phase rectiligne, on va leur faire aborder un virage qui les introduira sur une piste circulaire; ils prendront le virage tous ensemble, bien groupés au moyen d'une force exercée convenablement, et les voilà sur la piste circulaire. C'est la deuxième phase de l'accélération qui commence; on va alors les laisser exactement sur la même piste : **ils vont toujours décrire la même trajectoire, mais seront accélérés à chaque tour** ou plusieurs fois par tour. Le peloton ira de plus en plus vite. Naturellement, on inclinera de plus en plus la piste de façon à éviter le dérapage, et on essaiera également de toujours les grouper en peloton : il ne faut pas que ceux qui ont pris un peu de retard continuent à en prendre, mais il faut qu'ils rattrapent le peloton dans les

Là se place un passage difficile; c'est celui de la bonne introduction du peloton de particules sur la piste circulaire. Cette introduction pose des problèmes qui ont finalement été bien résolus, et le peloton s'engage dans le vide à l'intérieur de l'anneau. C'est précisément cet anneau qui fait une boucle de 800 m de circonférence, mais son propre diamètre intérieur n'est que de quelques dizaines de centimètres, si bien que les faisceaux doivent s'engager à la perfection si l'on veut éviter la dispersion des particules ou la rencontre des parois. Admettons que l'on ait réussi l'opération et que le peloton se propage maintenant à l'intérieur du tube; il faut qu'il suive parfaitement l'axe de l'anneau, car on veut lui faire faire beaucoup de révolutions. Mais, pour qu'il n'aille pas en ligne droite et qu'il se courbe selon le cercle principal de l'anneau, il faut que le peloton soit soumis à un **champ magnétique** qui seul sera capable de modifier sa trajectoire rectiligne en la transformant en une trajectoire circulaire. Là se situe une différence essentielle avec le cyclotron. **Dans le cyclotron, la trajectoire était une hélice et les particules dessinaient des arcs de cercle de rayons croissants. Ici, le cercle est toujours le même,** aussi le champ magnétique doit-il être parfaitement adapté à la vitesse de la particule : au début, le champ magnétique doit être faible pour ne pas trop courber les particules, encore assez peu rapides, et le champ doit progressivement, **en synchronisme parfait avec l'accélération des particules,** croître suivant

Phot. CERN, Presse-Sports.



L'accélérateur linéaire qui sert à obtenir l'accélération préliminaire en ligne droite des protons.

une certaine loi, de façon que le faisceau continue à suivre toujours l'intérieur de l'anneau. Nous avons vu que, dans le cyclotron, le champ magnétique est parfaitement constant : on n'a pas à s'en occuper (ou pas énormément). Ici, au contraire, le problème du synchronisme entre le champ magnétique et l'accélération des particules doit être résolu avec une précision inouïe.

C'est au cours de cette promenade circulaire que le peloton va se sentir accéléré plusieurs fois par tour. Ce sera naturellement par un champ électrique convenable, et il faudra évidemment avoir affaire à des champs électriques de haute fréquence, car les particules vont vite. Au bout d'un grand nombre de révolutions, l'énergie aura ainsi augmenté considérablement, et le champ magnétique sera parvenu au maximum qu'on peut obtenir : les particules du peloton auront toutes acquis l'énergie formidable de 25 milliards d'électron-volts.

Donnons quelques chiffres, assez étonnants d'ailleurs. Notre peloton va se composer non pas de quelques cyclistes, mais d'environ 400 milliards de protons à chaque injection; ces 400 milliards de protons doivent être conservés bien groupés tout le long de l'accélération, et le synchrotron devra veiller à ce que les protons en retard puissent revenir au sein du peloton. Il devra veiller également à ce que, partout où les nombreux effets qui peuvent perturber une trajectoire se produiront, il n'y ait pas risque pour le faisceau de venir heurter les parois. Quant au temps pendant lequel les particules vont être accélérées, il est à peine de plus d'une seconde; mais comme elles vont très vite, elles font de nombreux tours pendant ce temps, et on peut estimer à environ 300 000 le nombre de kilomètres qu'elles vont parcourir à l'intérieur du tube pendant le temps d'accélération. On voit qu'il s'agit là d'une énorme distance, ce qui montre à quel point les précautions qu'il va falloir prendre seront draconiennes. On conçoit qu'un tel appareil exige pour sa construction les meilleurs techniciens et les meilleurs physiciens de tout un continent.

Mais pourquoi donc le tube de 800 m est-il tellement étroit? Pourquoi prendre simplement 20 cm et non pas 1 m de diamètre intérieur? La raison en est fort simple : tout au long de l'anneau se trouvent de puissants aimants, qui forment un enveloppement quasi continu du tube. Ce sont ces aimants qui donnent le champ magnétique nécessaire pour courber les trajectoires, et ils l'établissent dans tout le volume occupé par l'anneau. Si ce volume est très grand, il faudra fournir un champ magnétique beaucoup plus important, et cela coûtera beaucoup plus cher. La production d'un champ magnétique intense dans le volume de l'anneau est extrêmement onéreuse en énergie électrique et en dispositifs de production du courant et du champ; on a le plus grand intérêt à réduire l'anneau autant que possible, tout en lui conservant sa longueur de 800 m, qui est une longueur obligatoire pour l'obtention de l'énergie de 25 milliards d'électron-volts que l'on envisage. On ne serait d'ailleurs pas arrivé à diminuer comme il convient le diamètre intérieur de l'anneau si l'on n'avait pas eu recours à des combinaisons d'aimants très particulières : l'appareil du CERN est le premier dispositif dans lequel on a groupé les aimants déviateurs d'une certaine façon, avec des gradients alternés, afin d'éviter, à l'intérieur même du tube, les oscillations du peloton, oscillations qui auraient exigé un diamètre plus grand avec les méthodes classiques.

Lorsqu'on a construit le CERN, on a donc pris un gros risque; celui de réussir ce dispositif de gradients alternés; le succès a couronné les efforts faits. Il faut dire qu'à la même époque les Russes entreprenaient un appareil du type classique, sans gradients alternés, qui fonctionne actuellement, mais qui est beaucoup plus lourd que celui du CERN et qui correspond à une énergie beaucoup moins grande. Au contraire, l'appareil américain AGS est tout à fait équivalent à celui du CERN et donne des résultats aussi satisfaisants : il fut mis en service quelques mois plus tard.

Avec le synchrotron, nous sommes en pleine relativité. Dès qu'un proton atteint une énergie de 1 à 2 milliards d'électron-volts, sa vitesse est déjà presque égale à celle de la lumière, et, entre 2 et 25 milliards d'électron-volts, sa vitesse n'augmente presque plus. Pendant toute la fin de l'accélération, la vitesse reste parfaitement constante, et l'énergie augmente pourtant. On peut se rendre compte que l'énergie augmente, car il faut un champ magnétique plus intense pour conserver les trajectoires sur le même cercle. On sait également que l'énergie augmente, par l'observation des phénomènes de désintégration et de matérialisation que l'on peut réaliser; tout cela se mesure à la perfection.

Voilà que nous avons fait connaissance maintenant avec les deux grands appareils de la physique nucléaire moderne, et nous allons retourner dans les halls, pour voir un peu ce qui se passe avec le peloton accéléré. Toutes les trois secondes, un signal caractéristique indique le départ des protons, qui arrivent sur leur cible un peu plus d'une seconde après leur départ. Après quoi, un intervalle de près de deux secondes est conservé pour la préparation de l'injection suivante : on peut dire que toutes les trois secondes 400 milliards de protons viennent frapper une cible. Nous avons déjà vu l'intérieur du hall en arrivant. Les cibles sont toutes petites, quelques millimètres ou quelques centimètres de matière au plus; les protons les traversent soit une fois, soit un grand nombre de fois lorsqu'ils ont atteint leur énergie maximale, et, à chacune de ces petites traversées, dont l'ensemble ne dure pas plus d'un millième de seconde, une énorme quantité de noyaux de la cible se trouvent soumis au bombardement des protons. On peut dire en gros qu'il y a plusieurs centaines de milliards de désintégrations nucléaires effectuées à partir de chaque injection. Naturellement, les noyaux vont se désintégrer en leurs constituants, si bien que beaucoup de protons et de neutrons vont partir avec des énergies de plusieurs milliards

Phot. CERN.

d'électron-volts dans toutes les directions, et spécialement vers l'avant. Mais il n'y a pas que cela, et le plus intéressant pour les physiciens est précisément d'étudier les particules instables que l'on peut créer au cours des interactions de grande énergie. Il sera question de ces particules plus loin, au chapitre de la physique corpusculaire, mais nous pouvons dire dès maintenant que l'on crée, à partir d'énergie, des mésons π chargés des deux signes, des mésons π neutres, des mésons lourds (ou mésons K) chargés positivement ou négativement, ou encore des mésons K neutres, ou même des nucléons excités sous forme de ce que l'on nomme les « hypérons ». Le problème de savoir ce qui se passe quand deux protons se rencontrent à grande énergie est pour les physiciens le problème fondamental; il est au centre de toutes leurs préoccupations; il exige l'exploration des régions profondes des noyaux et même des nucléons, qui en sont les constituants élémentaires.

Admettons donc cette énorme quantité de phénomènes; la question qui se pose est la suivante : comment peut-on les déceler? Tout cela se passe si vite et avec un amoncellement si brutal et abondant de particules de toute sorte émises dans toutes les directions, avec toutes les énergies, que l'on a peine à comprendre comment l'on peut s'y reconnaître et étudier certains effets de façon particulière et précise.

les antiprotons

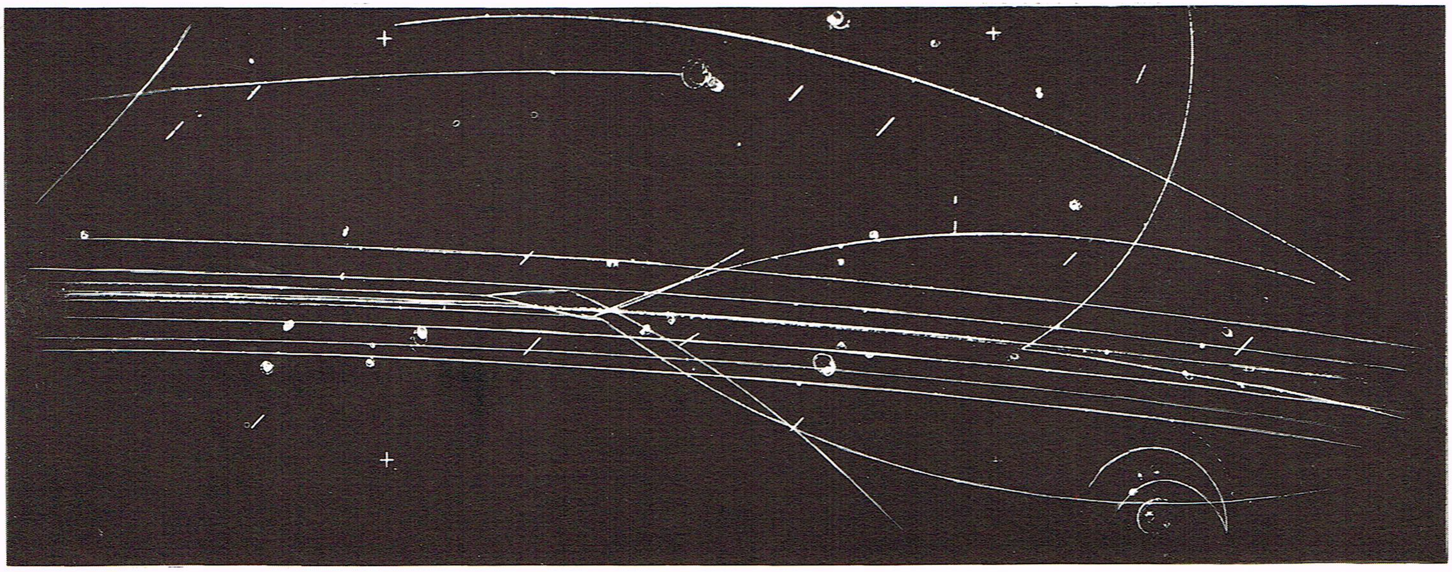
Il n'y a pas que les mésons et les hypérons qui soient créés; il y a encore les antiprotons et les antineutrons, et même les antihypérons. Nous sommes dans le domaine admirable de l'identité de la masse et de l'énergie : si l'on calcule, d'après la formule extrêmement simple qui relie l'énergie aux masses, celle qui correspond à la masse du proton, on trouve un milliard d'électron-volts environ. On peut donc penser qu'avec une énergie disponible d'un milliard d'électron-volts on pourra créer un proton : pourquoi, en effet, le proton échapperait-il à la loi d'Einstein? Mais on sait également qu'il est probablement difficile d'ajouter des protons à la nature et que le nombre de ces particules doit être constant dans l'univers. Il semble bien que ce soit une grande loi de la physique. En revanche, on sait également qu'il existe des antiparticules. On connaît déjà l'antiélectron ou électron positif, qui a — la charge électrique mise à part — les mêmes propriétés que l'électron; or, une antiparticule et une particule sont créées en même temps, et elles s'annihilent lorsqu'elles se retrouvent; à partir d'un rayon γ, on peut créer un électron et un antiélectron là où il n'y en avait pas, mais lorsque l'antiélectron rencontre un des nombreux électrons de la matière, les deux particules s'annihilent en redonnant des rayons γ.

De même, l'antiproton existe; c'est le noyau élémentaire de l'antimatière, celui de l'antihydrogène; c'est une particule analogue au proton, mais qui est chargée négativement; elle a les mêmes propriétés que le proton, à l'exception de celles qui correspondent à la charge électrique, puisque cette dernière est opposée. Du point de vue nucléaire, on peut concevoir des associations d'antiprotons et d'antineutrons pour former des antinoyaux, c'est-à-dire de l'antimatière sous une forme analogue à la matière.

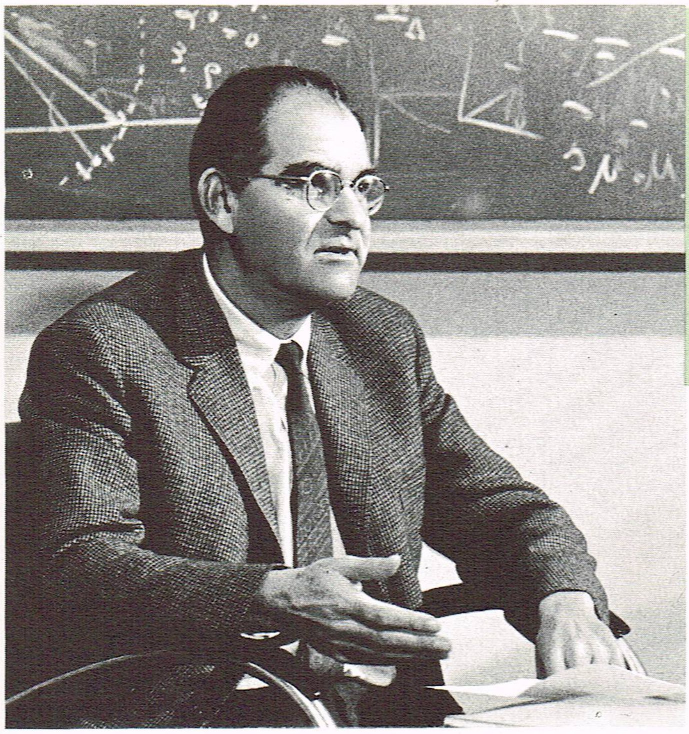
Si l'on crée un antiproton, on peut probablement le faire à partir d'énergie, si l'on dispose d'une énergie suffisante pour créer à la fois un proton et un antiproton. Un antiproton seul ne peut être créé, car ce serait ajouter une antiparticule lourde (baryonique) dans l'univers, ce qui ne semble pas concevable. En revanche, créer un proton et un antiproton ensemble est concevable, et c'est bien ce que l'on fait avec le synchrotron : on crée des couples « proton-antiproton ».

Que faire avec l'antiproton? Il faut d'abord, comme nous l'avons vu, séparer des quelques antiprotons les centaines de millions de mésons π, créés beaucoup plus facilement et qui accompagnent les antiprotons. On fera donc la séparation en sélectionnant l'ensemble des particules négatives d'une certaine énergie et en séparant, dans cette énergie, les antiprotons des autres. On aura ainsi réalisé ce faisceau que nous avons vu en entrant dans le grand hall. Les antiprotons vont interagir avec la matière, et l'on sait bien que si un antiproton s'arrête au voisinage d'un proton, les deux vont s'annihiler complètement. Que vont-ils faire en s'annihilant? Comment l'énergie va-t-elle être libérée? Il n'y aura plus ni proton ni antiproton, mais probablement émission de mésons ou d'associations plus compliquées, qui s'en iront un peu comme des toupies instables, qui auraient un certain moment de rotation et pourraient se séparer après en certaines composantes. Tout cela est extrêmement intéressant à étudier; c'est pour cela qu'on fait des faisceaux d'antiprotons et qu'on les reçoit dans les grandes chambres à bulles.

Trajectoires d'antiprotons rapides dans la chambre à bulles à hydrogène de 81 cm.
Cette photographie a permis de découvrir une antiparticule, l'« antixi », dont l'existence avait été prévue théoriquement.



Phot. CERN.



◀ **V. F. Weisskopf, directeur général du CERN de 1960 à 1966.**

▼ **J. B. Adams, responsable de la construction du synchrotron.**

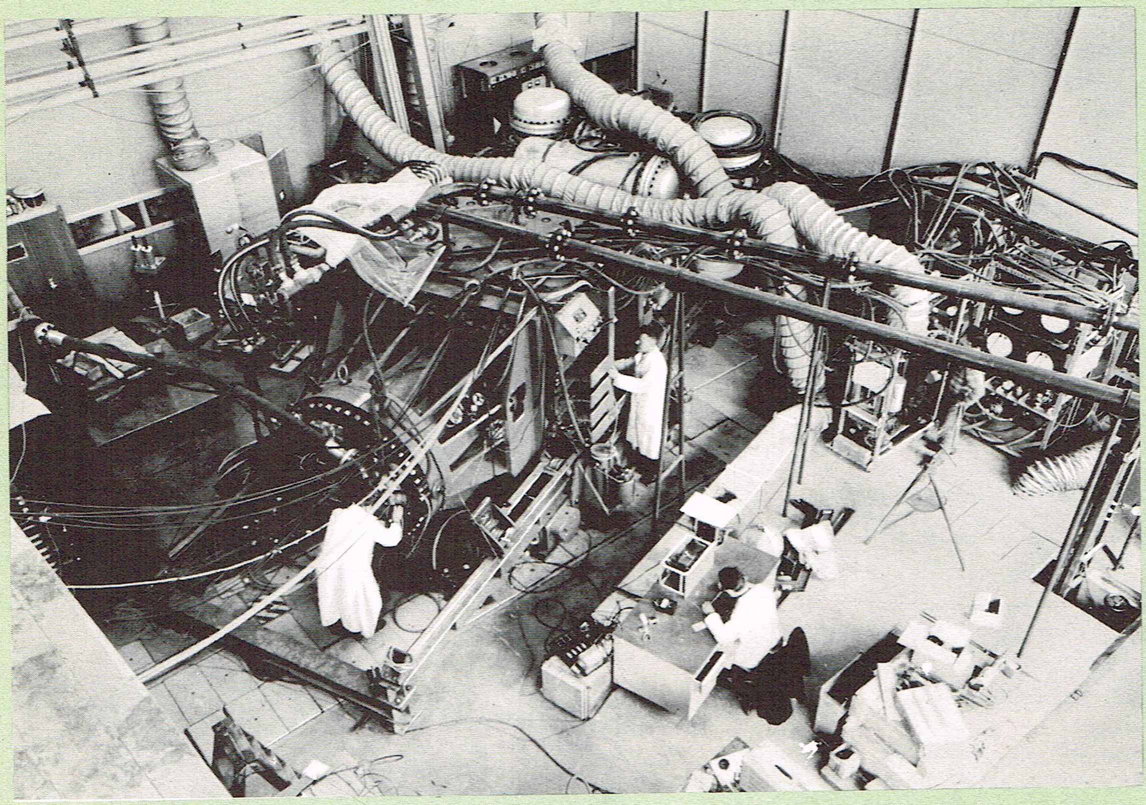


les grandes chambres à bulles

Nous voilà maintenant derrière de très gros blocs de béton, dans une petite casemate bien isolée, à plus de 100 m de l'anneau, à l'extrémité d'un faisceau fin d'antiprotons ou de mésons π. Dans cette casemate se trouve une chambre à bulles. C'est un appareil énorme, de plus de 100 tonnes, très mystérieux au premier abord par son caractère étrange et extraordinairement complexe. Si l'on veut le schématiser, on peut dire que la partie essentielle est une boîte transparente contenant un liquide. Ce n'est pas n'importe quel liquide, mais, pour étudier les interactions les plus simples, on mettra de l'hydrogène sous forme de liquide, c'est-à-dire de l'hydrogène liquide à une température très basse, environ 25 °K, c'est-à-dire 248 ° au-dessous de 0 °C. Donc, imaginons une boîte d'une grande capacité, une centaine de litres par exemple, contenant de l'hydrogène liquide sous une pression de quelques atmosphères. Naturellement, cela ne sera pas sans difficultés, attendu que tous les joints deviennent cassants à très basse température, que l'hydrogène liquide n'est pas de tout repos, que les précautions les plus grandes doivent être prises à cause des très réels dangers d'explosion présentés par une telle installation. Si l'on abaisse brusquement la pression dans le liquide, c'est-à-dire si, étant à 6 atmosphères, on ouvre une vanne pour tomber à la pression atmosphérique, alors l'hydrogène se met à bouillir; c'est ce qui se passe quand on abaisse suffisamment la pression d'un liquide : on sait bien que si l'on fait le vide au-dessus de l'eau ordinaire elle se met à bouillir. C'est cette ébullition qu'on utilise pour repérer les trajectoires des particules. N'en disons pas plus maintenant; on décrira la chambre à bulles plus tard, en montrant quels sont les phénomènes qui interviennent. Ce qui nous importe maintenant, c'est de savoir qu'avec un dispositif de ce genre on peut, tout au début de l'ébullition, repérer les trajectoires des particules électrisées en mouvement rapide qui traversent la chambre : si un proton ou un antiproton très rapide traverse une chambre à bulles au moment précis où l'on fait la détente, les petits centres électrisés qui ont été formés sur le trajet de la particule servent de germe pour les premières bulles. Il faut se dépêcher, photographier très vite, sinon tout se met à bouillir d'une façon très brutale, et on ne voit plus du tout de trajectoires. Cet instrument, inventé il y a fort peu d'années, a valu à Donald Glaser, jeune physicien de trente ans, le prix Nobel de physique, car la chambre à bulles est actuellement l'instrument le plus merveilleux pour observer les trajectoires de particules et avoir une vision parfaite des interactions les plus compliquées. On verra ultérieurement beaucoup d'exemples d'application de cette technique.

Il n'y a pas que les chambres à bulles, il y a également beaucoup d'instruments électroniques, car les rayons qui sont émis ou qui appartiennent au faisceau secondaire du synchrotron sont détectables également grâce à des compteurs d'électrons. Nous sommes déjà familiarisés avec les compteurs de Geiger, qui pétaradent lorsqu'on les approche d'une source radioactive, et qui servent à déceler la radio-activité. Mais il y a également les scintillateurs, qui sont une forme plus raffinée de compteurs d'électrons, et qui peuvent compter des millions de passages par seconde. Il y a également d'autres appareils extrêmement ingénieux, qui ont été imaginés par Cerenkov et qui permettent de savoir si une particule passant dans le dispositif a une vitesse supérieure ou inférieure à une certaine

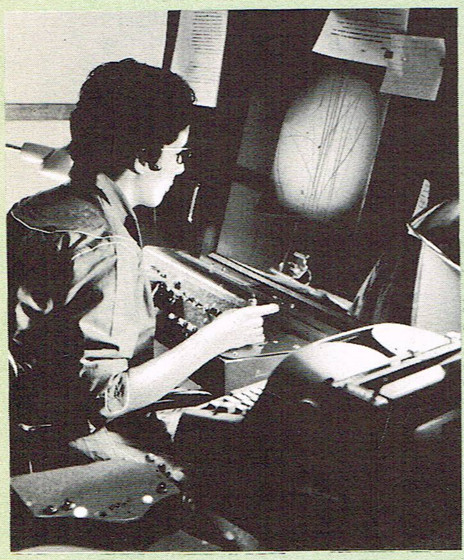
Phot. CERN.



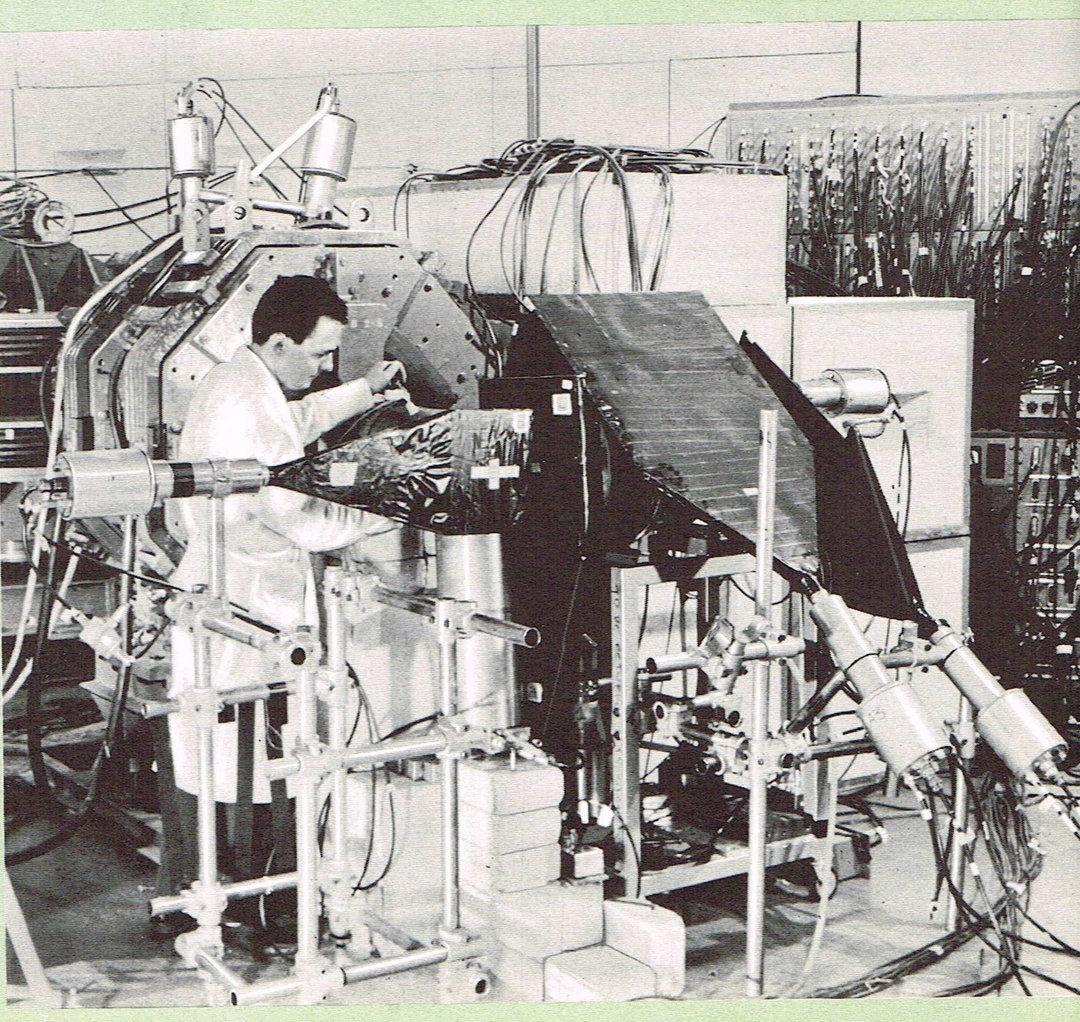
Vue d'ensemble de la chambre à bulles à liquide lourd.

Page ci-contre, le professeur Leprince-Ringuet observe les trajectoires apparaissant dans cet instrument.

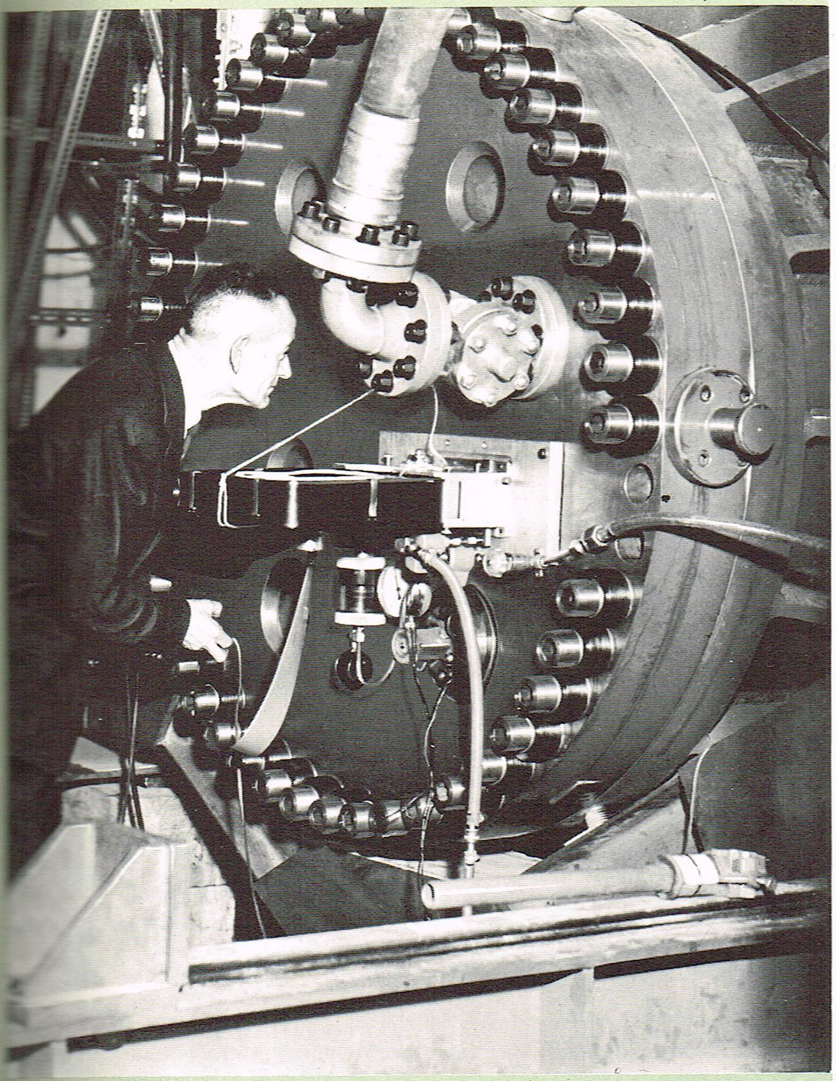
Ci-dessous, dispositif de détection électronique des particules.



« Scanning-girls », tel est le nom de ces aides-physiciennes chargées d'observer systématiquement les photographies pour en inventorier les événements.



Phot. CERN.



Phot. Ecole polytechnique.

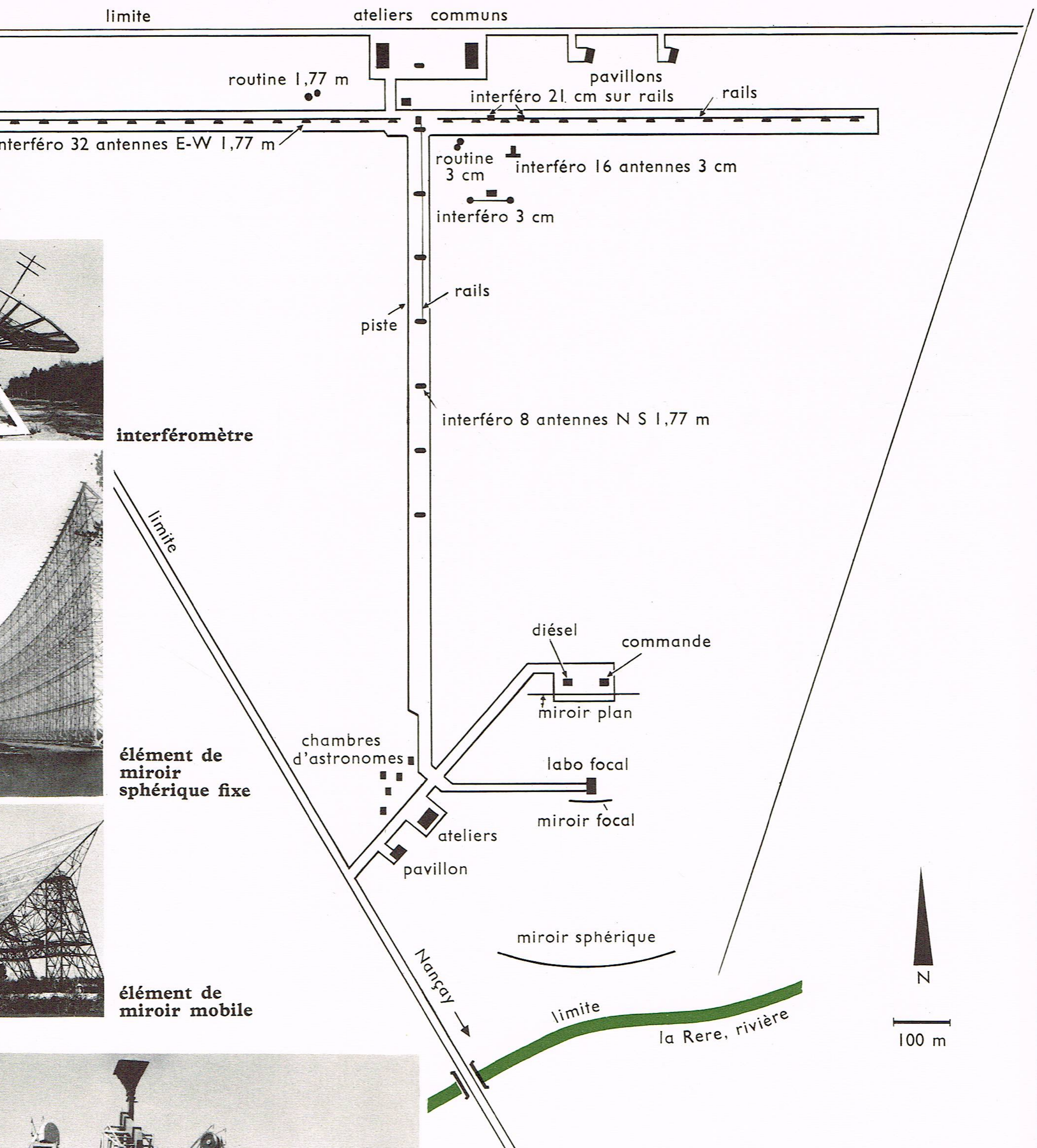
valeur. Tout cet ensemble connaît actuellement des développements considérables, et l'on est étonné par l'immense déploiement d'électronique dans toutes les pièces.

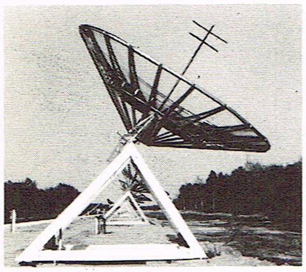
Et maintenant quittons les accélérateurs de particules et promenons-nous dans le reste du CERN. Nous y voyons des centaines de techniciens et de physiciens soit théoriciens, soit expérimentateurs; les uns préparent les expériences, les autres discutent sur certains résultats. Il y a beaucoup de réunions communes, appelées « séminaires », où l'on étudie une question particulière, par exemple la production des mésons K dans les annihilations des antiprotons. Toute une partie du CERN est consacrée à l'exploitation des photographies, car une expérience fournit au moins 100 000 photographies analogues à celles qui sont présentées dans cet ouvrage, et chacune d'elles n'est pas seule — elle est accompagnée d'au moins deux autres vues stéréoscopiques du même événement —, si bien que l'on doit faire une étude systématique de centaines de milliers de photographies. Il est bien évident qu'on ne peut pas le faire sans de très grandes machines électroniques; aussi, le CERN possède-t-il toute une armée de jeunes gens qui ont pour tâche d'explorer les photographies et de les introduire dans un dispositif qui transforme l'information des traces photographiques en information sur cartes perforées ou sur bandes magnétiques : un événement intéressant est ainsi transformé en une trentaine de cartes perforées; si l'on veut étudier 100 000 événements, on voit qu'il faut beaucoup d'organisation et de méthode pour classer toutes ces cartes. Après cela, il va falloir faire ce qu'on appelle un « programme », c'est-à-dire expliquer à une puissante machine, un ordinateur électronique I. B. M. ou Bull, ce qu'on désire qu'elle fasse : il lui faut commencer par lire la première trace à partir de son entrée dans la chambre jusqu'à l'endroit où elle s'arrête, donner les indications de son rayon de courbure; il faut ensuite passer à la trace n^0 2, donner les informations sur sa direction, son orientation, son changement de courbure si elle n'a pas toujours le même rayon, etc.; puis passer à l'origine de deux autres traces qu'on voit un peu à côté, et, finalement, fournir tous les renseignements sur toutes les traces qui nous intéressent. Fabriquer un tel programme a nécessité à Berkeley six physiciens et six techniciens pendant deux ans. Il ne s'agit pourtant que d'écrire des instructions à envoyer à la machine électronique, et qu'elle soit capable de digérer et de suivre. Il y a là de beaux problèmes d'électronique, de mémoires, de programmation, d'information dont nous voudrions bien avoir la clef.

Nous quittons le CERN émerveillés par une technique tellement évoluée. Nous avons envie de connaître la relativité, la masse et l'énergie, l'électromagnétisme, les transmutations nucléaires, les constituants fondamentaux de la matière, ces particules instables que l'on crée et qui s'évanouissent, l'antimatière, et également aussi toutes ces subtilités de l'électronique qui sont indispensables actuellement dans la vie moderne.

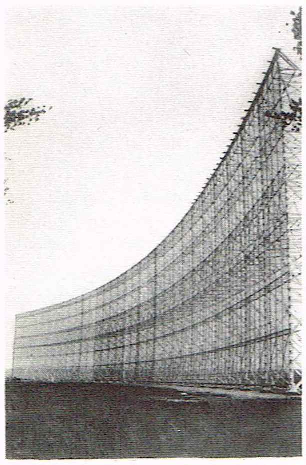
Louis **LEPRINCE-RINGUET**

Plan général de la station radio-astronomique de Nançay.

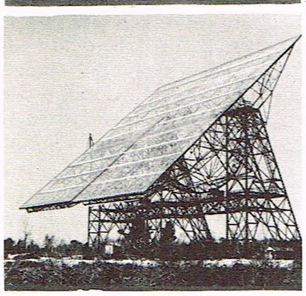




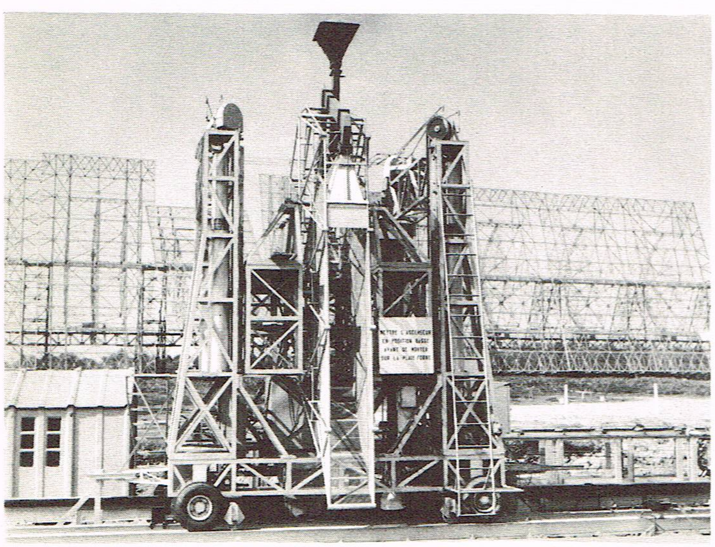
interféromètre



élément de miroir sphérique fixe



élément de miroir mobile



Le « chariot focal », contenant les récepteurs radio, comprend, au milieu, une petite antenne en cornet qui collecte l'énergie incidente envoyée par le miroir sphérique. Au fond, le plan mobile. Le chariot focal est situé à 230 m du réflecteur sphérique.

Phot. Observatoire de Paris.

l'observatoire radio-astronomique de NANÇAY

Quittons maintenant la frontière franco-suisse et transportons-nous par la pensée au centre même de la France, en Sologne, où nous attend un autre instrument mondialement célèbre, voué au culte de l'infiniment grand, et non plus de l'atome.

Près de cent miroirs, répartis sur un kilomètre carré de lande solognote : Nançay est un monde réservé à la science; c'est un monde à zéro, une ou deux dimensions. Un monde sans dimension, comme la Sologne elle-même semble l'être au promeneur englouti dans sa forêt; sa vue se perd à quelques mètres dans la mêlée des bouleaux et des pins, des genêts et des bruyères. Nul horizon lointain ne vient inviter à rêver; le ciel lui-même est à peine visible à travers les branchages; tout est confiné dans un microcosme de senteurs et de bruissements d'où s'évadent, au gré des caches, un lapin, un faisan, tout à coup matérialisés, *ex nihilo*. Nançay est aussi un monde à une dimension, quand, au sortir d'un fourré, sans transition, le regard embrasse la longue ligne des trente-deux miroirs de l'interféromètre est-ouest. Deux points de fuite à l'horizon, un pan de ciel, un souffle d'air; on débouche sur un monde différent. Et la nuit, lorsque Jupiter vient par-dessus les cimes des conifères refléter sa lumière impassible, presque éternelle, dans les eaux des marécages, une impression puissante d'ère secondaire envahit l'esprit. Devant la quasi-permanence des astres et de la nature, les corolles métalliques des miroirs semblent être des fleurs ouvertes pour une seule nuit, toutes dirigées vers le même point des cieux, un point où l'œil ne voit que quelques faibles étoiles; mais un point qui certainement les attire toutes, un point d'où parvient un évanescent message d'ondes invisibles, qui, plus loin, dans une cabane, anime un récepteur. Tout à l'heure, dans les bois, l'homme sentait son univers ramené à un point; ici, il sent son existence ramenée à un bref instant de la durée. Ce n'est qu'en arrivant sur l'aire du grand miroir qu'il se rassure. Ici, une vaste zone a été défrichée; en une légère ondulation, elle domine la Rère. Au loin, on aperçoit un clocher, quelques fermes, une maison de campagne. Des pistes, des canaux de drainage, des câbles sillonnent la large lande. D'innombrables repères géodésiques entourés de barrières la piquettent. Et surtout, majestueux et élancés, se répondant par-dessus les bruyères, le miroir plan et le miroir sphérique se font un face-à-face parfait, encore accentué par les cabanes servantes disposées à leurs pieds. Nançay est ici un monde à deux dimensions, largement ouvert sur l'univers. La grandeur de l'entreprise, le fait qu'elle soit encore en expansion, les millions de francs qu'elle a déjà coûté, la promesse de découvertes nouvelles sur la nature du Cosmos, tout concourt à mettre l'esprit dans un état de dynamisme euphorique; un dynamisme tel d'ailleurs que, lorsque le visiteur s'enquiert de l'évolution future d'un observatoire de radio-astronomie, l'astronome, s'envolant délibérément selon la troisième dimension, déclare que le prochain « Nançay » sera construit sur la Lune.

ce que représente Nançay

Nançay est un exemple typique de grand observatoire radio-astronomique. De par le monde, il en existe au plus une dizaine : un aux Pays-Bas, deux en Grande-Bretagne, un en Australie, trois aux Etats-Unis, deux en U. R. S. S. Tous ont des projets grandioses, à l'étude ou en voie de réalisation. L'instrument le plus audacieux était un miroir parabolique de 180 m de diamètre, orientable dans toutes les directions. (A vrai dire, cet instrument n'était pas destiné à une utilisation purement scientifique; en 1962, l'armée américaine a dû en arrêter la construction.) Nançay vient au premier rang dans la taille des instruments en cours de montage, avec son miroir mesurant 300 m sur 35 m. Il suffit donc de se laisser guider dans une visite de cette station pour obtenir un panorama général de la radio-astronomie.

la radio-astronomie, science jeune

On le sait, cette science est très jeune, puisqu'elle date de la Seconde Guerre mondiale, la technologie des ingénieurs du radar en ayant inspiré les réalisations. L'avènement de récepteurs sensibles pour ondes métriques, décimétriques et même centimétriques, ondes que l'ionosphère de la Terre n'arrête pas, allait permettre d'exploiter la fenêtre ainsi ouverte sur le monde extérieur. La découverte en fut fortuite cependant, et bien antérieure à la période d'exploitation, puisque c'est l'Américain Janski, en 1931, qui, en étudiant les parasites atmosphériques, observa un maximum de réception lorsque la Voie lactée passait au zénith; mais sa découverte tomba dans l'oubli.

La radio-astronomie n'est pas une science différente de l'astronomie; elle en est une partie. Une partie nouvelle et importante. Elle doit sa particularité au fait que l'ionosphère est transparente aux ondes électromagnétiques de longueurs comprises entre 1 cm et 30 m, et que, dans ce domaine, les récepteurs sont évidemment bien différents des récepteurs optiques. Avec la création de récepteurs extra-terrestres, les deux domaines se rejoindront. Pour l'instant, la collaboration étroite des spécialistes des deux branches est un fait, et aussi la clef d'une compréhension plus rapide et plus profonde des phénomènes. C'est là un facteur important de progrès.

L'aspect des astres est très différent vu en ondes optiques ou en ondes radio, comme on peut le concevoir lorsqu'on constate la grande différence qui existe déjà entre une photographie en rouge et une photographie en bleu d'une même région du ciel, à tel point même que certaines des radiosources les plus intenses (tel Bootes A) sont des objets optiquement presque invisibles, même au plus grand télescope optique, alors que Vénus, si brillante, n'émet guère d'ondes radio.

L'émission radio-électrique des astres revêt pour beaucoup un aspect mystérieux et inquiétant. Du subconscient surgit l'image d'un nouveau Branly qui, de Mars, essaierait d'entrer en contact avec la Terre. Il n'en est rien, bien sûr, quoique, à un stade plus avancé de la technique, il sera possible d'essayer de capter les émissions intelligentes que, éventuellement, des êtres évolués peuvent émettre dans l'espace. L'émission radio-électrique des astres doit plutôt faire penser aux parasites émis par un orage ou aux rayons calorifiques émis par une pièce de fer chauffée, ou à la raie jaune du sodium, ou, encore, à la lumière émise par le faisceau des synchrotrons. Ces quatre procédés d'émission sont effectivement à l'œuvre dans l'univers, comme nous le verrons plus tard.

un radiotélescope

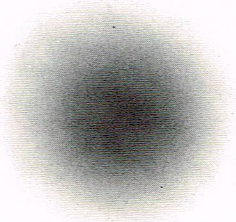
Pour l'instant, il suffit de réaliser que l'aspect d'un radiotélescope est déterminé par la longueur d'onde utilisée. L'exemple le plus simple est donné par un miroir d'observation solaire qui fonctionne sur 1,77 m de longueur d'onde. La surface réfléchissante est un grillage grossier de 5 m de diamètre; il suffit, en effet, pour qu'elle soit considérée comme parfaite, que ses défauts soient nettement plus petits que la longueur d'onde; la situation est la même qu'en optique : la surface d'un miroir de télescope nous semble parfaite, mais, à une bactérie, elle semblerait être un affreux terrain rocailleux.

Si l'on braque un tel radiotélescope vers une source ponctuelle, l'énergie reçue par l'antenne sera maximale lorsque l'axe passera exactement par la source. Cependant, si le télescope n'est pas pointé exactement, l'antenne recevra encore de l'énergie de la source; du fait que la longueur d'onde utilisée n'est pas minuscule par rapport au diamètre du miroir, de forts effets de diffraction vont intervenir. Ce sujet de la diffraction, si important en optique, sera traité plus tard; qu'il nous suffise

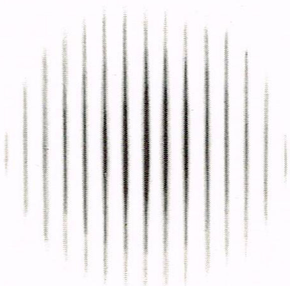


Télescope sur ondes métriques : un miroir parabolique de 5 m de diamètre concentre les rayons sur un dipôle placé au foyer; un câble coaxial transporte les potentiels induits à un amplificateur; l'ensemble du télescope est fixé sur une monture azimutale pour viser le Soleil lors de son passage au méridien.

Carte d'une radiosource ponctuelle faite à l'aide...



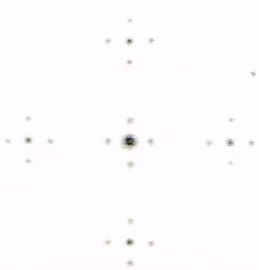
... d'un miroir simple



... d'un interféromètre à deux antennes



... d'un interféromètre à plusieurs antennes



... d'un interféromètre en croix

Phot. Observatoire de Paris.

de dire que, si l'on reporte sur une carte céleste l'énergie reçue d'une source ponctuelle en pointant le radiotélescope dans des directions variées au voisinage de la source, on verra que cette source a l'aspect d'une tache floue, comme le montre la figure ci-contre; pour le miroir d'observation solaire, cette tache a un diamètre de 10°; deux sources célestes éloignées de 10° l'une de l'autre lui sembleront fondues en une tache oblongue; si elles n'étaient qu'à 1°, elles seraient indiscernables l'une de l'autre. 10° est ici ce qu'on appelle le « pouvoir de résolution » de l'instrument. La radio-astronomie est très défavorisée par rapport à l'optique au point de vue de la résolution des instruments; pour qu'un radiotélescope fonctionnant sur ondes de 1 m ait le pouvoir de résolution de l'œil, soit 1/100 de degré, il faudrait qu'il ait 100 m de diamètre; et pour avoir le pouvoir de résolution du télescope du Mont-Palomar, il faudrait qu'il ait 100 km.

Quant à l'énergie recueillie par le radiotélescope, elle sera proportionnelle à la surface de l'instrument. On comprend maintenant l'intérêt qu'il y a à posséder des instruments de grande dimension; malheureusement, des impératifs technologiques et économiques limitent les possibilités, et des compromis doivent être trouvés.

l'interféromètre

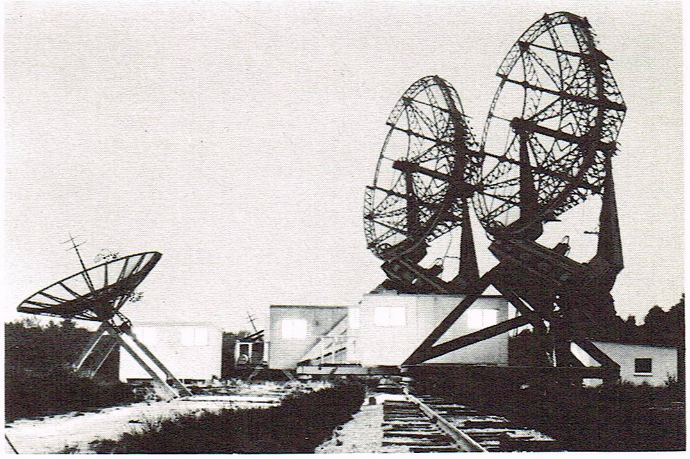
C'est ainsi que, si l'on veut obtenir une grande résolution pour des sources intenses, on peut opérer par interférométrie avec deux miroirs éloignés l'un de l'autre, tels ceux que représente la photographie. Les signaux reçus par les deux antennes sont mélangés dans le récepteur. Il en résulte qu'une source située dans le plan de symétrie de l'appareil enverra ses ondes simultanément sur les deux antennes; les amplitudes s'ajoutant, la réponse sera maximale. Si, en revanche, la source se trouve un peu en dehors du plan de symétrie, les ondes arriveront en se contrariant et donneront une réponse plus faible, ou même nulle. On a affaire au phénomène d'interférence, qui sera décrit en détail dans le tome II. Disons brièvement, comme tout à l'heure, que la carte d'une source ponctuelle serait encore une tache floue, mais striée de franges parallèles régulièrement espacées.

On peut même, en plaçant les deux antennes à des distances variées les unes des autres, arriver à reconstituer l'aspect réel d'une région céleste; l'inconvénient de cette méthode est la nécessité de faire des mesures successives. On peut y remédier en augmentant le nombre des antennes travaillant simultanément; on obtient alors un instrument tel que cet interféromètre à seize miroirs. Les signaux venant de chaque antenne sont mélangés dans le récepteur; le résultat final est que, parmi les franges dues aux deux miroirs extrêmes, il n'en reste pratiquement qu'une toutes les huit, les autres étant très affaiblies. Remarquons que cet instrument étant destiné à fonctionner sur 3 cm de longueur d'onde, la structure des miroirs est beaucoup plus précise; le grillage est remplacé par de la tôle, et le dipôle par un cornet. Chaque fois que le soleil passe au travers d'une frange, sa surface est explorée tranche par tranche et donne des enregistrements tels que celui de la photographie de la page suivante, où l'on voit le disque solaire sur lequel se superpose une tache émissive.

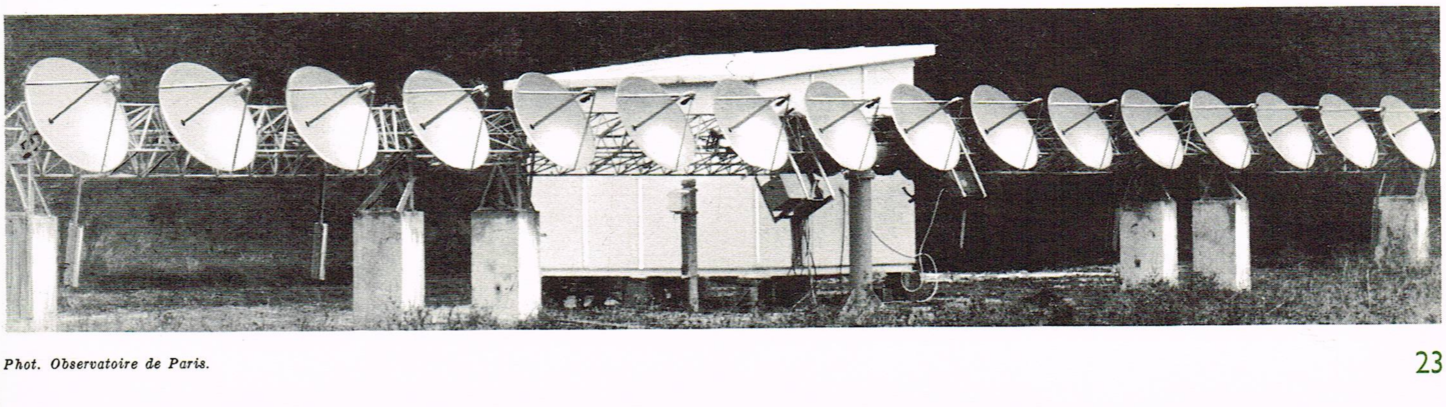
On peut encore améliorer l'instrument en lui adjoignant une autre ligne d'antennes disposée perpendiculairement à la première, comme le montre la photographie de la page suivante. La carte d'une source ponctuelle obtenue avec un tel interféromètre en croix est composée d'un réseau rectangulaire de taches disposées à l'intérieur de la tache floue.

Des méthodes ingénieuses de combinaison des signaux permettent d'exploiter au maximum l'information reçue par ces interféromètres.

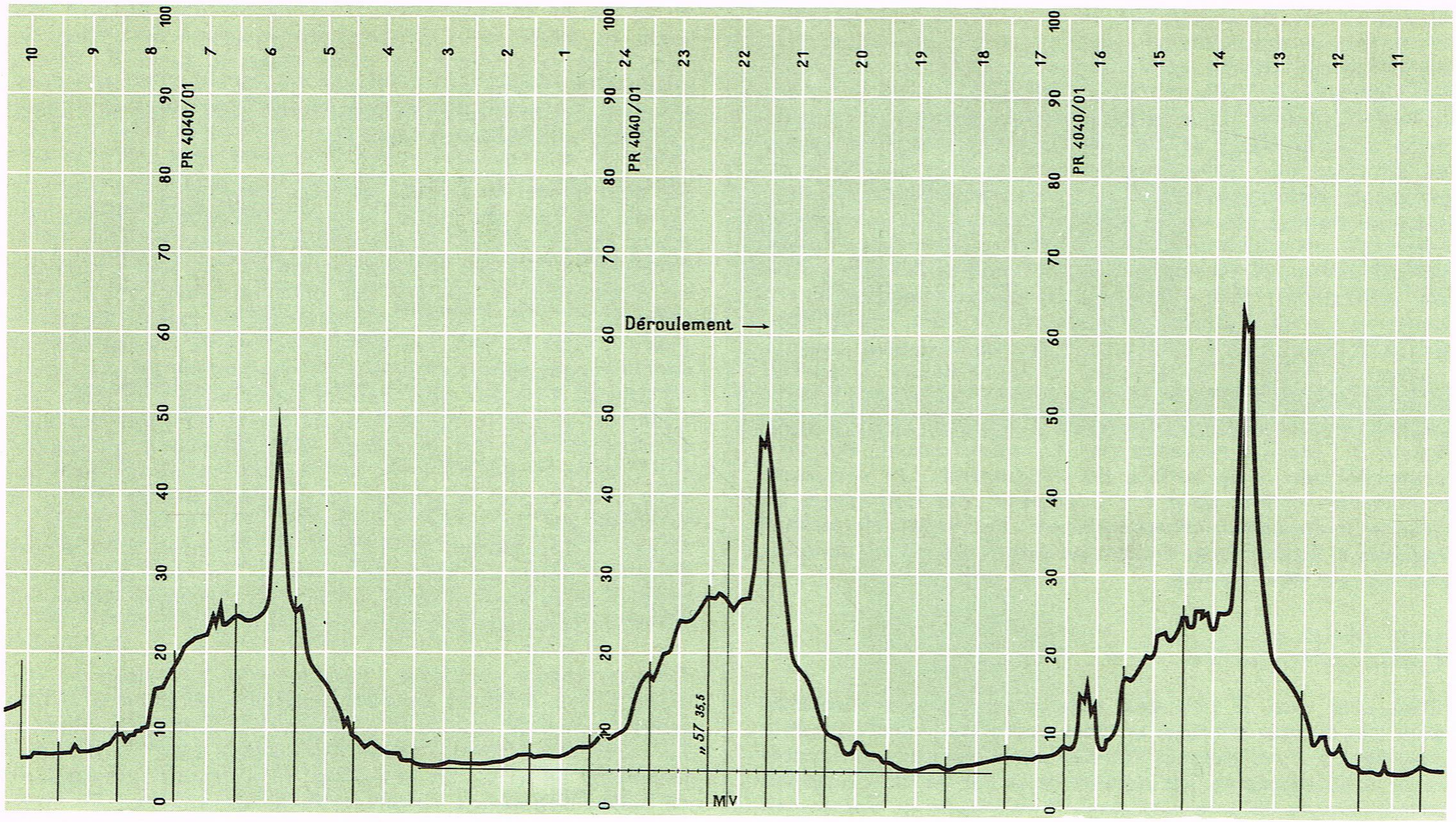
Interféromètre sur ondes décimétriques à deux miroirs mobiles sur rails. Chaque miroir, monté en équatorial, peut suivre l'astre dans son mouvement diurne. Chacun transporte une cabane où sont logés les amplificateurs. Les signaux sont ensuite envoyés dans la cabane sur le sol, à gauche, où ils sont combinés pour fournir les franges d'interférence.
A l'extrême gauche se trouve un des éléments de l'interféromètre sur ondes métriques.



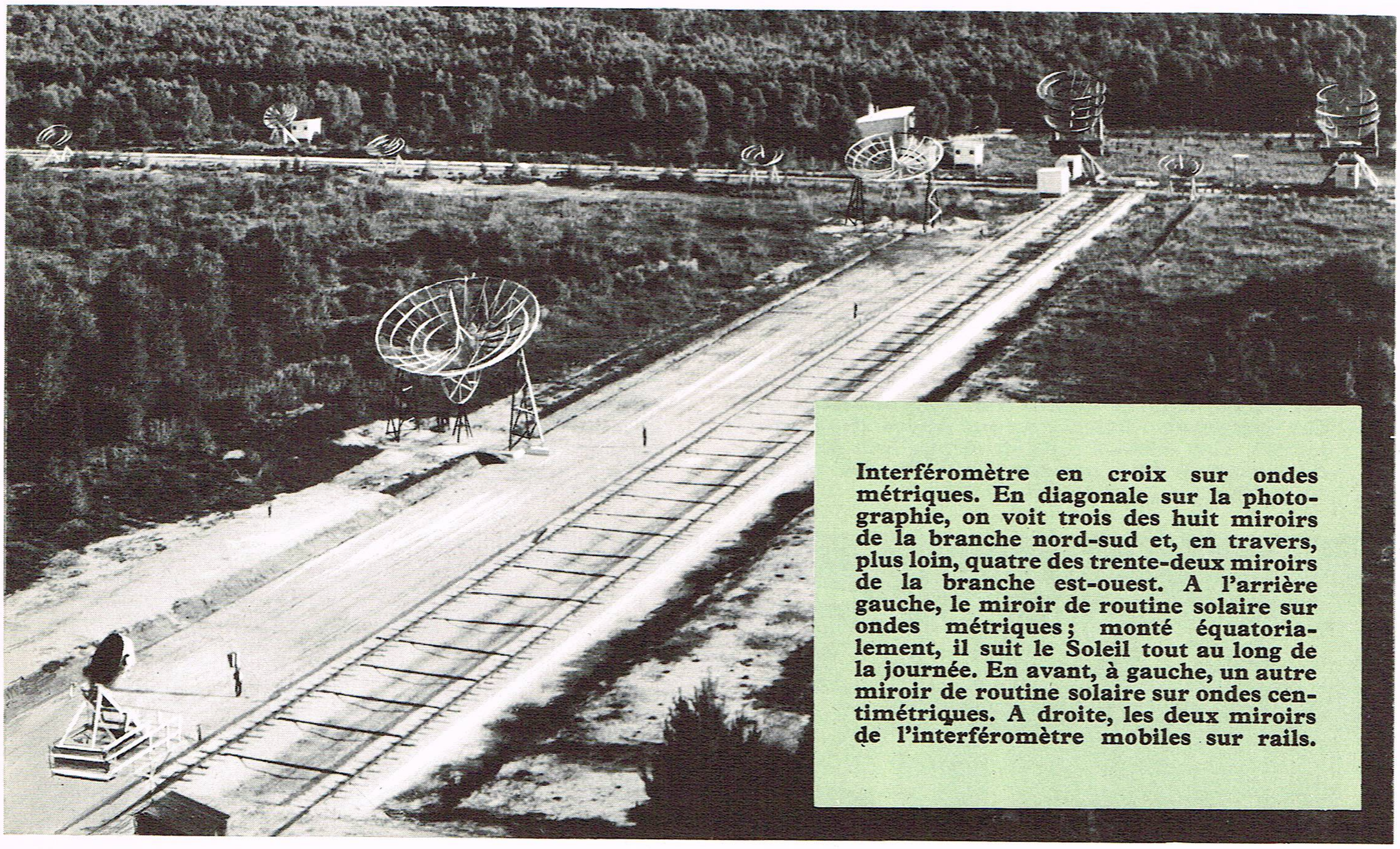
Interféromètre est-ouest à seize miroirs sur ondes centimétriques. Les seize miroirs sont montés sur une poutre est-ouest orientable en azimut. Les ondes, captées par des cornets, sont envoyées par des guides dans la cabane abritant les récepteurs. Cet instrument fonctionne sur ondes de 3 centimètres. Ces miroirs sont en tôle et leur structure est beaucoup plus précise que celle des miroirs en grillage.



Phot. Observatoire de Paris.



Passage du Soleil au travers de trois franges de l'interféromètre sur ondes métriques à trente-deux antennes ; chaque minute, un trait vertical est tracé par l'enregistreur. Le Soleil met quatre minutes pour traverser une frange ; il est donc, en ondes métriques, deux fois plus large que le Soleil optique. Remarquer la présence d'une tache émissive étroite et intense sur un des bords du Soleil.

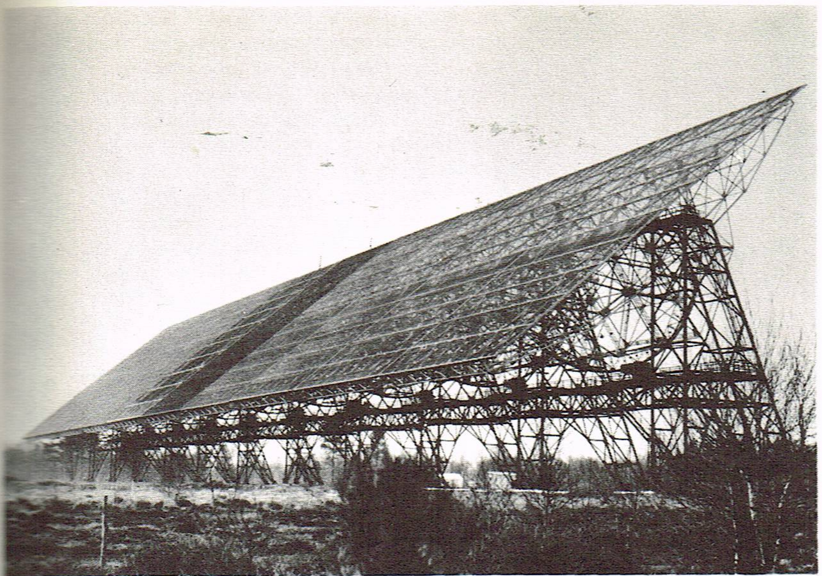


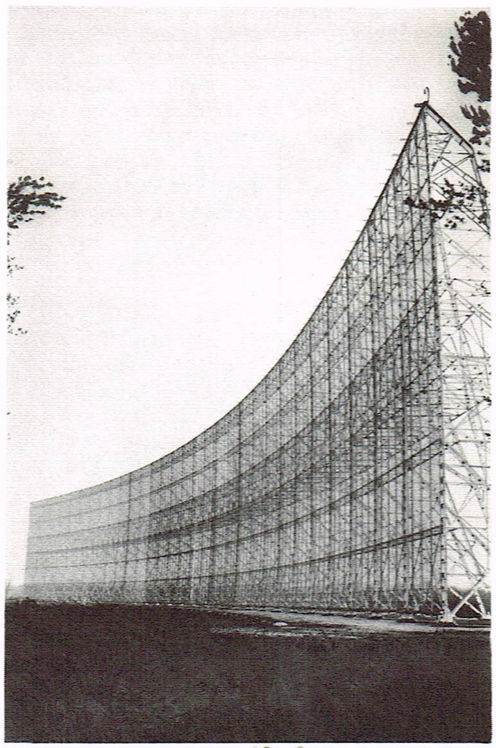
Interféromètre en croix sur ondes métriques. En diagonale sur la photographie, on voit trois des huit miroirs de la branche nord-sud et, en travers, plus loin, quatre des trente-deux miroirs de la branche est-ouest. A l'arrière gauche, le miroir de routine solaire sur ondes métriques ; monté équatorialement, il suit le Soleil tout au long de la journée. En avant, à gauche, un autre miroir de routine solaire sur ondes centimétriques. A droite, les deux miroirs de l'interféromètre mobiles sur rails.

Phot. Observatoire de Paris.

le grand miroir

Mais, bien sûr, au point de vue de la sensibilité, ces instruments ne peuvent donner plus que ce que pourrait détecter une antenne dont la surface totale serait égale à la somme des surfaces des antennes composantes, et, si l'on veut d'emblée la structure d'une radiosource, sans aucune ambiguïté, le grand miroir à surface continue est la seule solution pratique. Là encore des compromis sont nécessaires : une grande taille n'est réalisable qu'au prix d'une réduction de la mobilité de l'appareil. Le cas extrême actuellement est celui du miroir sphérique de Porto Rico, d'un diamètre de 300 m, qui est monté à même le sol dans une cuvette naturelle aménagée (à vrai dire, il est destiné à des études de radar-astronomie, et, en déplaçant légèrement l'antenne réceptrice au voisinage du foyer, il est possible, grâce au mouvement diurne de la Terre, d'explorer une petite zone du ciel). Un compromis a été réalisé pour le grand miroir de Nançay, instrument qui ne peut se déplacer que selon un axe et qui n'a de grandes dimensions que selon la direction horizontale. Il est réduit à ne pouvoir observer les astres qu'au méridien du lieu, et n'a une grande définition angulaire que selon la direction est-ouest. Mais sa conception originale vaut qu'on s'y arrête. Il consiste essentiellement en un miroir plan mobile autour d'un axe horizontal dirigé selon la direction est-ouest, qui renvoie les faisceaux venant d'un astre situé au méridien sur un miroir fixe sphérique. Ce dernier concentre les rayons en son foyer (photographies ci-contre). Cet instrument est destiné surtout à faire des études sur la raie de 21 cm, mais la grande précision avec laquelle il a été réalisé permettra de l'utiliser sur des longueurs d'onde plus courtes (11 cm, et même 6 cm). Le miroir plan est composé de dix panneaux de 20 m sur 40 m, mobiles autour d'un axe horizontal, et tendus de grillage dont les mailles sont de l'ordre du centimètre et dont la planéité est réalisée à quelques millimètres près. A environ 500 m en face de lui se trouve le miroir sphérique fixe de 35 m de haut et de 560 m de rayon de courbure; le foyer se trouve entre les deux miroirs, et consiste en une embouchure de 5 m de hauteur, qui concentre en définitive les rayons sur l'appareil récepteur très sensible placé au fond d'un guide d'ondes. Tout près de l'antenne réceptrice se trouve un laboratoire dans lequel est placé l'amplificateur. L'ensemble du bâtiment et du miroir parabolique focal est mobile le long d'un rail, pour permettre la mise au point d'un instrument, comme pour une longue-vue ordinaire.



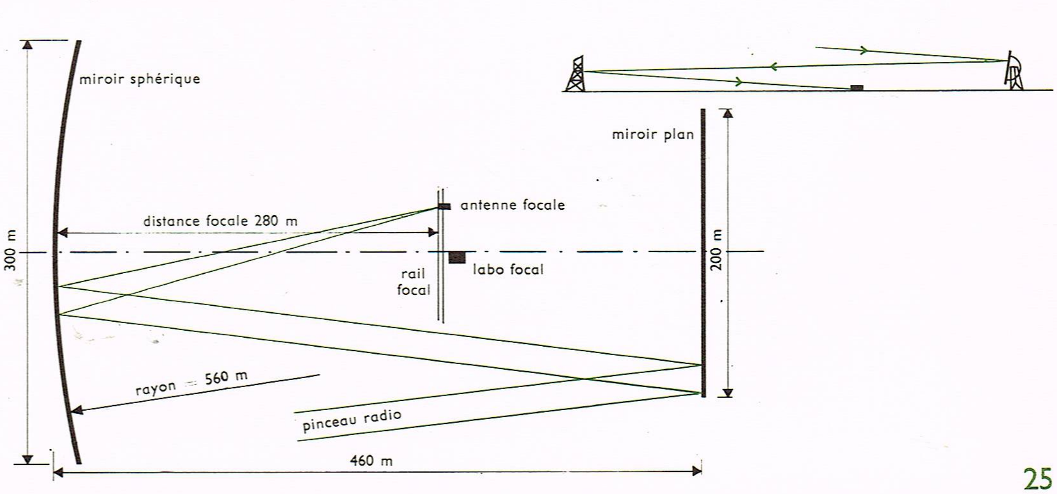


Ci-dessus, **miroir plan mobile du grand radiotélescope. On voit ici les dix panneaux de 20 × 40 m mobiles autour d'un axe est-ouest, construits en 1964.**

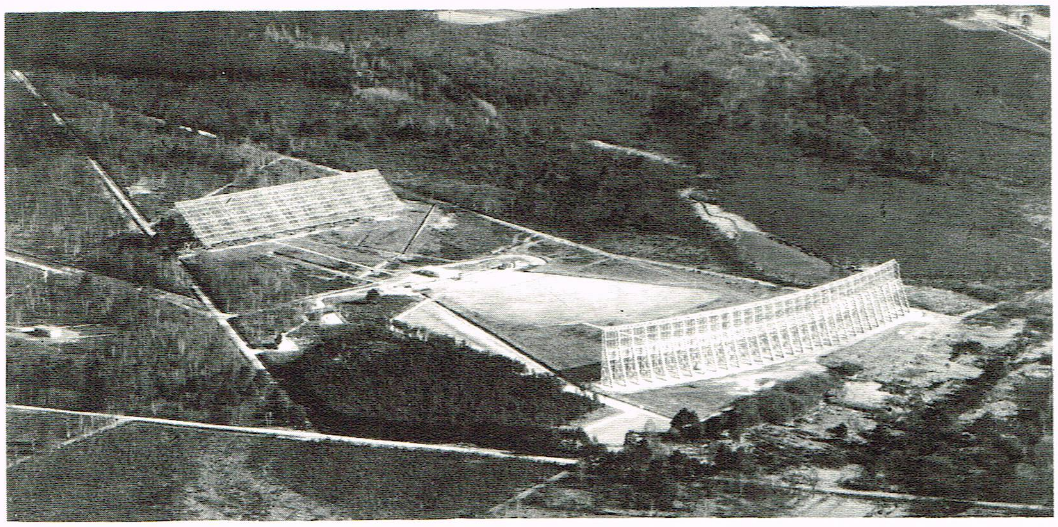
Miroir sphérique fixe du grand radiotélescope. La partie visible ici, achevée en 1965, de 35 m de haut et de 300 m de large, constitue l'ensemble du miroir définitif.

Schéma d'ensemble du grand miroir.

(V., page suivante, le miroir focal du grand radiotélescope.)



Phot. Observatoire de Paris.



Vue aérienne du grand radiotélescope de Nançay.

le récepteur

Les récepteurs utilisés en radio-astronomie sont parmi les plus sensibles qui soient. Pour fixer tout de suite les ordres de grandeur, sachons que la radio-galaxie la plus intense n'envoie qu'un millionième de microwatt sur un miroir de 300 m de diamètre dans un intervalle de fréquence de 10 MHz. Pour dégager d'une façon utilisable des signaux venant de sources encore plus faibles, il faut amplifier tellement que le bruit de fond de l'appareil devient une gêne majeure; la situation est équivalente à celle du radio-amateur qui veut capter des messages lointains et faibles au milieu de forts parasites. Ici, le bruit de fond provient surtout des mouvements aléatoires des électrons dans les parties résistives des circuits haute fréquence de tête. Actuellement, des études importantes sont en cours de par le monde (récepteur type Maser ou récepteurs dits « paramétriques ») pour réduire ce bruit de fond, inévitable, comme nous le verrons dans un prochain chapitre.

Le flux émis par la portion d'une radiosource vue dans l'antenne est mesuré en watt par mètre carré et par hertz. On l'exprime aussi par la température que devrait avoir un corps noir englobant entièrement l'antenne pour donner la même déviation sur l'enregistreur. Cette température est dite « température d'antenne »; elle est tout simplement proportionnelle au flux. Elle n'est pour le radio-astronome qu'un moyen commode et conventionnel de chiffrer ses observations; son origine provient de la technologie radio-électrique.

L'amplitude des fluctuations, bien visibles sur les enregistrements où l'amplification est poussée au maximum, est d'autant plus petite que la largeur de bande du récepteur (domaine de longueurs d'ondes admises par le récepteur) et la constante de temps de l'appareil sont grandes.

Cependant, la largeur de bande ne peut dans la pratique dépasser quelques dizaines de mégahertz. Aussi voit-on l'intérêt qu'il y a à augmenter le temps de pose : la théorie de l'information montre que le seul moyen d'améliorer la situation est d'augmenter le temps de mesure.

C'est pour cette raison que le grand miroir de Nançay a été construit de telle sorte qu'il peut suivre une source pendant une heure, lors de son passage au méridien, par un déplacement de l'antenne au foyer. Il est ainsi possible d'obtenir de bonnes images jusqu'à des distances angulaires du plan méridien de l'ordre de près de 10 °.

Ce long temps de pose est d'autant plus nécessaire au grand miroir qu'il est destiné principalement à des études sur la raie 21 cm de l'hydrogène atomique. En mesurant la fréquence exacte à laquelle on reçoit la raie d'un certain astre, on peut, par la mesure de l'effet Doppler-Fizeau, connaître la vitesse radiale de cet astre; on est donc conduit à faire des études dans des bandes étroites, de quelques centaines de kilohertz, ou même de quelques kilohertz si les circonstances sont favorables. Et, pour des bandes aussi étroites, le bruit de fond est très grand.

les observations

Que voit-on? Qu'entend-on même? nous demande-t-on souvent. En fait, si l'on branchait un haut-parleur à la sortie du récepteur, on n'entendrait que le bruit caractéristique de « friture » des récepteurs poussés. Que voit-on alors? On ne voit rien d'autre en radio-astronomie que le trait d'une plume agitée sur une bande de papier, ou des trous enfoncés par une perforatrice dans des cartes. C'est un peu dommage du point de vue de l'esthétique de l'astronomie, mais il faut dire que, dans le domaine optique, les images sont aussi de plus en plus délaissées; c'est ainsi que, pour étudier des galaxies extrêmement faibles, on doit avoir recours aux techniques photo-électriques. Cependant, ces enregistrements arides permettent de reconstituer des cartes du ciel, tel qu'il serait vu par un œil sensible aux longueurs d'ondes radio. En opérant des balayages à des déclinaisons rapprochées, il est possible de reconstituer l'aspect de vastes régions du ciel. Voici, à titre d'exemple, page ci-contre, la carte du Petit Nuage de Magellan obtenue par des astronomes australiens sur la longueur d'onde de 21 cm et, en comparaison, une photographie de ce même nuage.

Cette carte constitue un exemple déjà évolué de travail radio-astronomique. Beaucoup d'observatoires sont encore engagés dans le travail fondamental de dénombrement de radiosources et d'établissement de catalogues où sont consignées leurs positions, même lorsque celles ci-sont relativement imprécises.

Le travail primitif de comptage de sources en fonction de leur intensité est pourtant déjà d'une grande importance du point de vue de la cosmologie. On en est un peu au stade de Herschel lorsqu'il comptait le nombre d'étoiles visibles par degré carré en fonction de leur éclat, quand on essaie de découvrir la structure de notre île stellaire, la Galaxie. Ce stade est aussi celui de l'astronomie extra-galactique optique, qui actuellement

Phot. Base d'Avord.

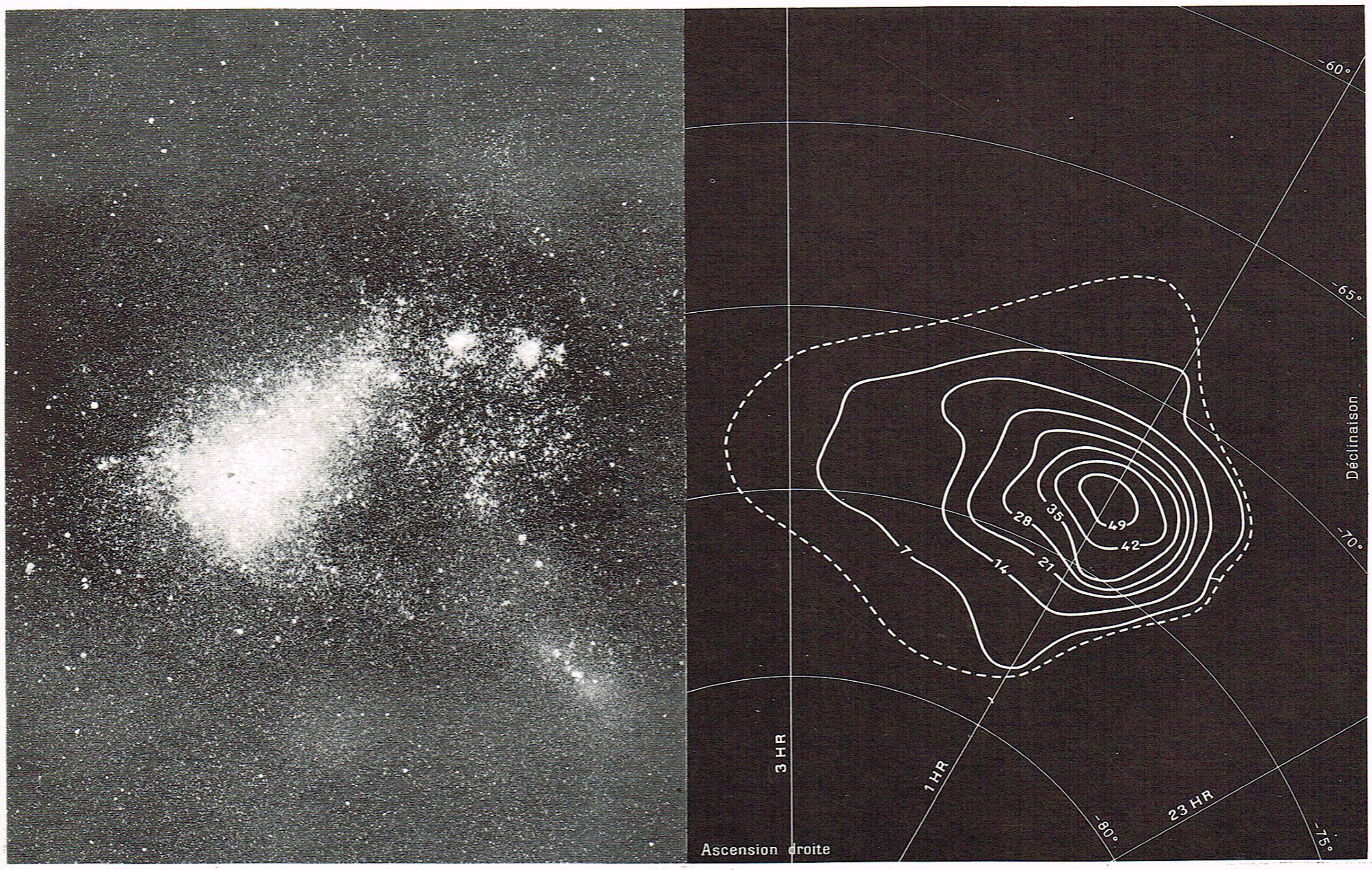
poursuit les comptages de galaxies en fonction de leur magnitude apparente pour essayer de découvrir la structure de l'univers dans son ensemble. En radio-astronomie, ce travail a une importance considérable, car les radiogalaxies sont telles que la portée des radiotélescopes en profondeur dans l'espace est au moins dix fois plus grande que celle des télescopes optiques (nous reviendrons sur ce point dans le tome II). Ce travail de catalogage est aussi très important pour résoudre le problème de l'identification des sources radio-astronomiques, car c'est en comparant avec l'objet optique, au cas où celui-ci est visible, que de grands progrès pourront être réalisés dans la compréhension de la nature des radiosources. Ces catalogues tiennent aussi lieu de répertoires de recherche pour les travaux plus minutieux effectués dans certains domaines, par exemple pour des mesures de positions plus exactes d'objets particulièrement intenses ou intéressants. Actuellement, les mesures de position les plus précises sont faites grâce à l'emploi de grands interféromètres, avec une précision de la seconde d'arc.

le soleil, radiosource vivante

Dans le domaine du système solaire, les mesures prennent un aspect différent, car il est possible d'y observer effectivement des variations au cours du temps, tandis que dans le domaine stellaire tout est statique à l'échelle de l'homme (il faut dire cependant que, depuis quelques années, on a mis en évidence des variations très importantes dans le flux radio d'une classe de sources tout à fait extraordinaires : les *radiosources quasi stellaires*. Certaines varient de 50 p. 100 en quelques mois. Parmi les astres du système solaire, c'est le Soleil qui est la source, à ce point de vue, la plus intéressante, d'abord par l'énorme intensité de son rayonnement radio, et aussi par les variations importantes dont il est le siège. Certaines de ces variations ont même un caractère tellement explosif que c'est la lenteur relative de l'instrument d'enregistrement qui limite la vitesse d'évolution des phénomènes mesurés. Un énorme matériel expérimental est accumulé sur le Soleil dans tous les domaines de fréquence radio aussi bien que dans le domaine optique, et le Soleil peut être considéré comme le laboratoire le plus proche où les réactions dans le plasma peuvent être étudiées. Jupiter aussi, mais à une bien moindre échelle, est le siège de variations importantes d'émissions radio, qui sont très probablement induites par l'activité solaire dans son ionosphère. Enfin, des variations sont également observées sous forme de scintillations dans les observations de haute précision des radiosources, que ce soit leur intensité qui varie ou leur position. Ainsi les radio-étoiles scintillent comme des étoiles visibles à l'œil.

Nous avons décrit les principaux appareils. Avant d'approfondir les diverses notions rencontrées au cours de cette brève

Photographie du Petit Nuage de Magellan et carte radiotélescopique de ce nuage en longueur d'onde 21 cm. Sur le réseau des coordonnées célestes (ascension droite et déclinaison) sont portées les quantités d'hydrogène atomique détectées; le nuage ressort clairement. Le Grand et le Petit Nuage de Magellan sont deux petites galaxies, satellites de la nôtre. Très proches, ils sont d'un intérêt exceptionnel pour l'observateur. Des centaines de radiosources y ont été détectées.



Phot. Commonwealth Scientific and Industrial Research Organization.

excursion, avant de comprendre quel puissant instrument d'étude du Cosmos est la radio-astronomie, avant de voir dans les chapitres qui vont suivre comment se révèle à nous l'univers, posons-nous une dernière question.

Comment s'effectuent ces observations d'aspect si varié? Il ne faut pas croire qu'il suffit au radio-astronome d'arriver au radiotélescope, de le pointer vers l'astre désiré, de faire l'enregistrement et de retourner dans son bureau l'étudier à loisir. Comme cela est le cas dans bien des domaines de la recherche scientifique, la technique de la radio-astronomie est poussée par l'astronome, avide de renseignements sur l'univers, au maximum de ce qu'elle peut donner. Telle radiosource qui serait mesurable en cinq minutes avec le grand miroir de Nançay ne présente plus un intérêt considérable du fait qu'elle a été déjà étudiée par des instruments moins puissants, mais poussés à la limite extrême de leurs possibilités; c'est donc vers de nouveaux objets, faibles par conséquent, que le grand radiotélescope se dirige, et, pour ceux-ci, ce n'est pas un ou même dix enregistrements qu'il faudra obtenir, mais quelquefois une centaine. Les travaux de réduction d'un si grand nombre d'enregistrements sont longs et demandent de nombreux ajustements, par exemple à cause de la variation de sensibilité du récepteur, du mouvement de l'axe du télescope dû au balancement de l'axe de rotation de la Terre du fait de la précession, de la nutation, etc. Ce n'est qu'après ce long travail qu'un résultat peut être dégagé. De même, la préparation des mesures demande beaucoup de travail, car il ne faut pas gâcher le temps d'un grand radiotélescope. Aussi faut-il choisir à l'avance les portions du ciel qu'il faudra viser afin d'en retirer le maximum de renseignements. Un travail particulier sur un astre donné peut donc demander quelques mois, et les grands travaux de cartographie, de catalogage, eux, peuvent demander plusieurs années.

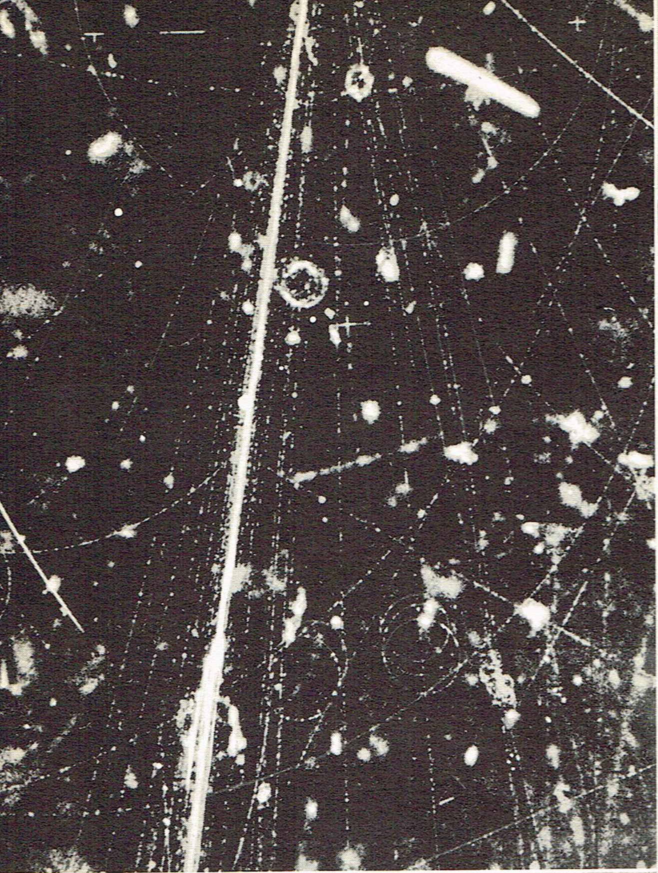
Pour utiliser au maximum un instrument de cette taille, toute une équipe de radio-astronomes s'y consacrera, chacun se spécialisant dans tel ou tel domaine. Les programmes doivent donc être organisés plusieurs mois à l'avance, afin d'obtenir un rendement maximal, du point de vue scientifique et horaire à la fois, surtout pour un instrument méridien où les sources passent à des heures déterminées dans l'axe de visée. Il arrive souvent que les sources intéressant deux ou plusieurs astronomes passent à la même heure devant le miroir; il faut alors faire un choix, qui, à Nançay, résulte de la discussion libre et amicale du groupe tout entier. Les recherches modernes nécessitent des instruments importants, coûteux, difficiles à mettre en œuvre; mais nous pensons que l'organisation, par trop lourde, qui serait nécessaire pour obtenir le fonctionnement complètement rationnel d'un groupe important de chercheurs, est un écueil qu'il faut éviter. Nous avons le sentiment qu'une organisation trop poussée de l'équipe risque de stériliser l'esprit de recherche, et qu'une certaine touche d'artisanat, d'individualité, de liberté, d'apparence de temps perdu même, doit teinter les grands groupes de recherche.

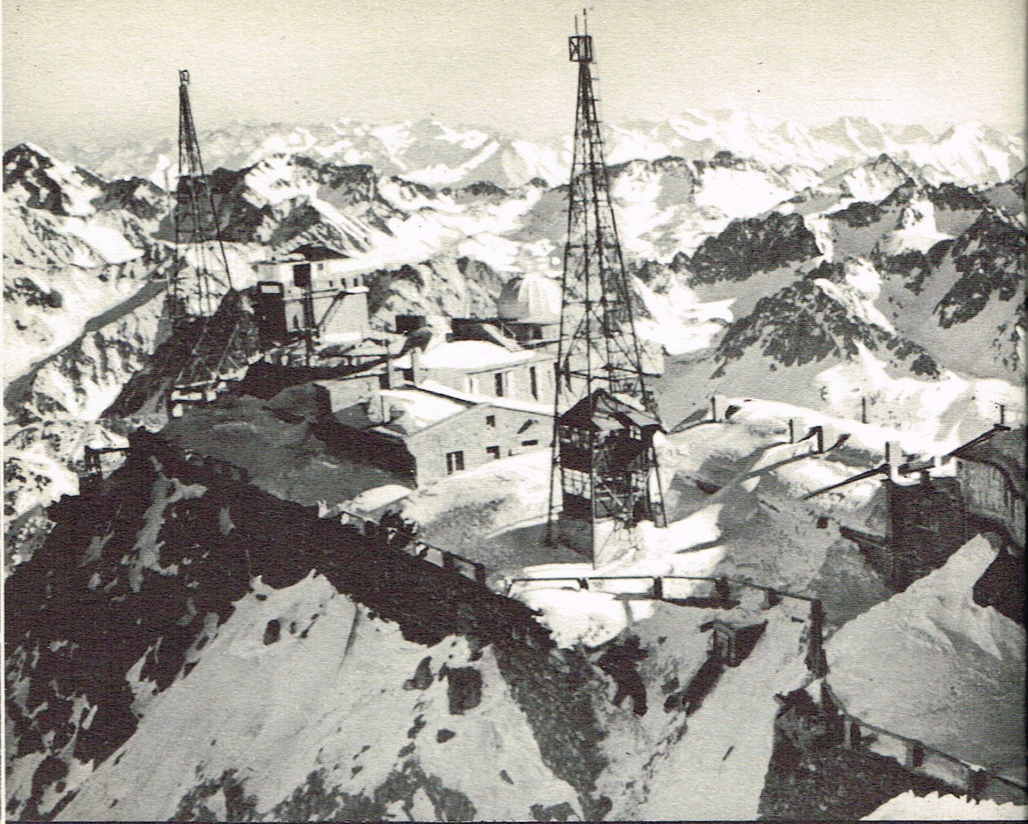
Combien coûte le grand radiotélescope de Nançay? Dix millions de francs. Cette somme semble bien dérisoire quand on essaie de la transposer en son équivalent en nombre d'heures de guerre, comparaison classique à laquelle pourtant il faudrait toujours revenir lorsqu'il s'agit de chiffrer le coût de réalisations destinées à enrichir le patrimoine intellectuel, et, en définitive, matériel, de l'humanité tout entière. Et pourtant ce prix peut sembler élevé à celui qui verrait le miroir de Nançay comme un grand panneau d'affichage routier de 1 000 t. Le coût de ce grand panneau vient de ce que tous les points de sa surface sont positionnés à quelques millimètres près, et qu'ils doivent le rester au cours des ans.

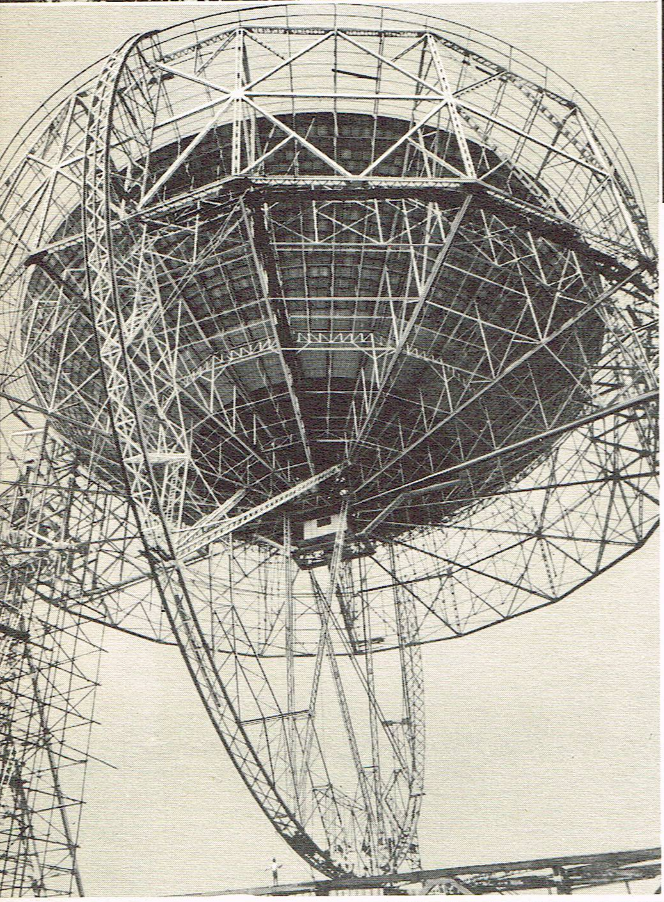
Jean HEIDMANN



Aériens de la branche est-ouest de l'interféromètre qui, à Nançay, assure la localisation et la poursuite des satellites artificiels. *Doc. C. S. F.*







En haut, **gerbe produite par l'impact d'un rayon cosmique et observatoire du pic du Midi de Bigorre.** *Ci-dessus*, **radiotélescope de Jodrell Bank (Grande-Bretagne).** *Page 31*, **le quatrième lancement expérimental d'une fusée « Saturn I ».**

l'exploration

Jusqu'à une époque récente encore (1961), l'homme a vécu rivé au sol de sa planète, ce qui ne l'a pas empêché de créer ces deux édifices splendides que sont la mécanique classique et la physique des ondes, de couronner le tout par la synthèse relativiste, puis d'élaborer, à partir des années 1920-1930, la physique moderne, d'essence probabiliste.

Mais d'autres perspectives s'offraient à lui : il pouvait capter et lire les messages précieux que lui adressait le Cosmos même, dans son immensité.

Et d'abord la lumière des étoiles, si faible mais si puissante en tant que moyen d'investigation. La contemplation du ciel fit rêver les premiers hommes, leur révéla la beauté et la grandeur de l'univers — monde aux apparences tellement simples et harmonieuses que les notions de permanence et de loi purent s'imposer à l'esprit humain : c'est en ce sens que l'on a pu parler de « l'astronomie, libératrice de la pensée humaine », et plaindre les Vénusiens, si, toutefois, ils existent, n'ayant d'autres perspectives qu'un ciel bas de nuages en mouvement désordonné. Mais la découverte de l'analyse spectrale décuplait la richesse de l'information reçue et permettait, grâce à une extrapolation aussi hardie que raisonnable sur la permanence des lois de la nature, de mieux connaître le centre des étoiles que le « sial », couche de l'écorce terrestre dont ne nous séparent pourtant que quelques kilomètres de roches. L'investigation humaine, après avoir franchi les frontières de la Galaxie grâce aux télescopes géants mis tour à tour en service, atteignait allégrement les limites de l'accessible, révélait des mondes morts depuis des milliards d'années et s'attaquait avec succès aux mystères mêmes de l'origine de l'Univers et de la Vie.

Les rayons cosmiques, aux énergies si considérables, révélaient l'intensité des champs électromagnétiques qui meublent le « vide interstellaire ».



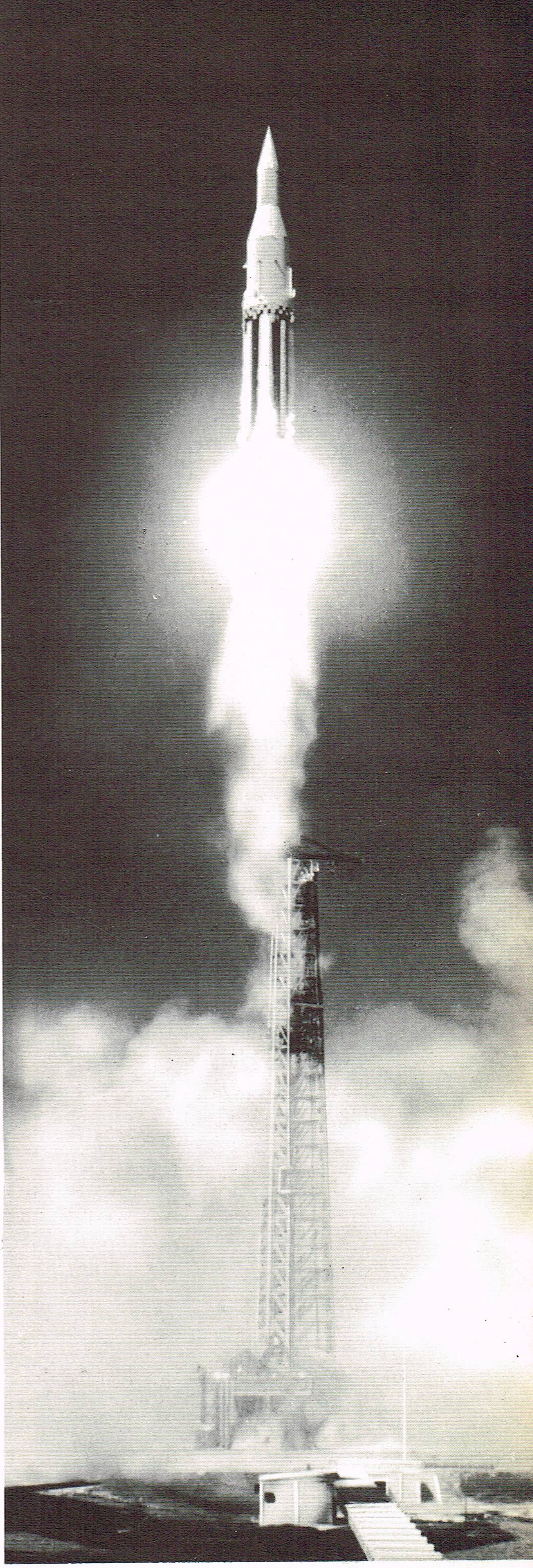
du Cosmos

Dernière-née, la radio-astronomie explorait le même univers sous d'autres longueurs d'onde et complétait notre connaissance de la Galaxie et du Cosmos. Enfin, le recensement des météorites confirmait l'unité chimique du monde et posait, en termes prometteurs, le problème de l'existence d'innombrables planètes habitables, peut-être même habitées.

L'homme ne s'est pas contenté de recevoir passivement ces messages, passablement déformés, appauvris et affaiblis par leur passage à travers l'atmosphère terrestre. Il s'est acharné, au prix des plus coûteux efforts, à franchir cette barrière, et s'est lancé à la conquête de l'espace, braquant sur la Lune et les autres planètes le faisceau puissamment concentré d'un « laser » et, surtout, l'arsenal de plus en plus complexe de ses fusées.

De cette dernière conquête, le bilan se révèle déjà positif, malgré le coût fantastique de la moindre réalisation : il apparaît clairement que les plus belles moissons ne peuvent être que le fruit d'une collaboration internationale, sincère et complète (CERN, Année géophysique internationale, Années internationales du Soleil calme, Observatoire de l'hémisphère Sud, exploration spatiale, lutte des biologistes contre le cancer, recherches sur le fonctionnement du cerveau, etc.), seule capable de faire reculer les frontières de la puissance et de la connaissance humaines. C'est probablement même par ce dernier trait que l'aventure scientifique influencera le plus fortement, et le plus bénéfiquement, l'aujourd'hui — dangereux mais exaltant — de notre civilisation.

Puissent les hommes prendre modèle sur cette réelle Communauté mondiale que forment maintenant les équipes de scientifiques, unies par la pensée, le langage, le but et la nécessité reconnue d'une collaboration réellement vécue dans l'espace et dans le temps.



Doc. Ecole polytechnique, U. S. I. S., C. C. A., phot. Yan.

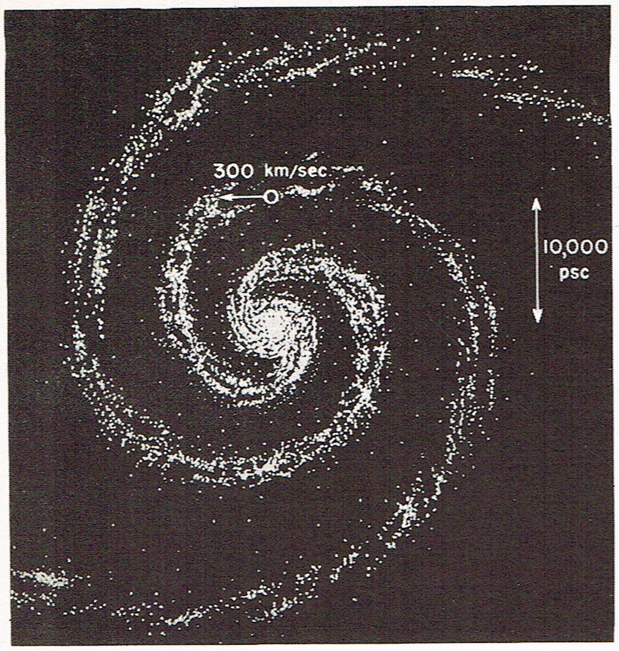
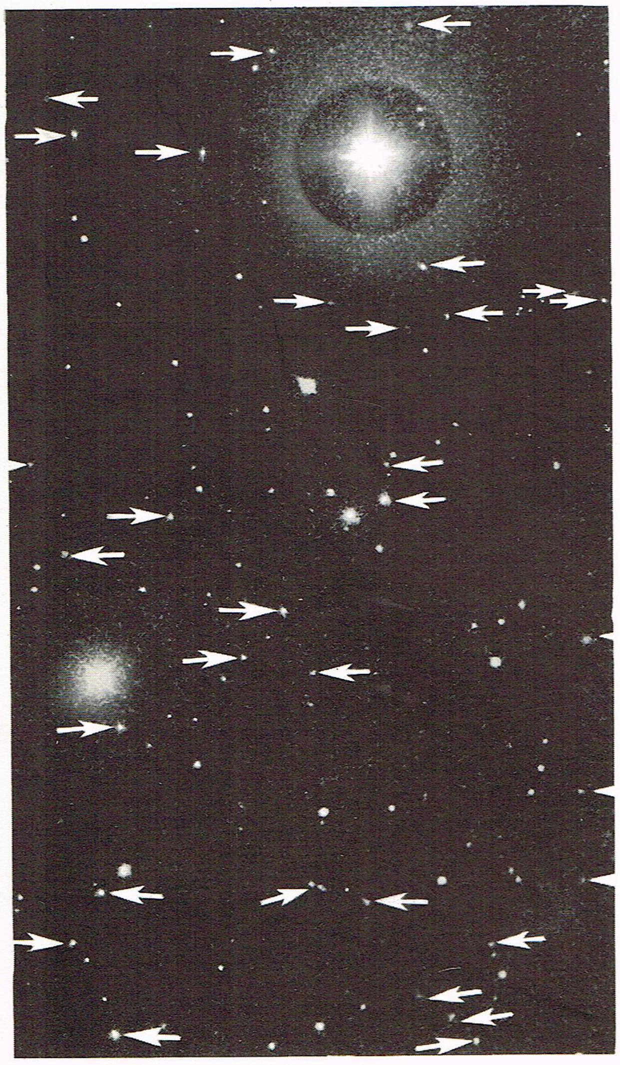
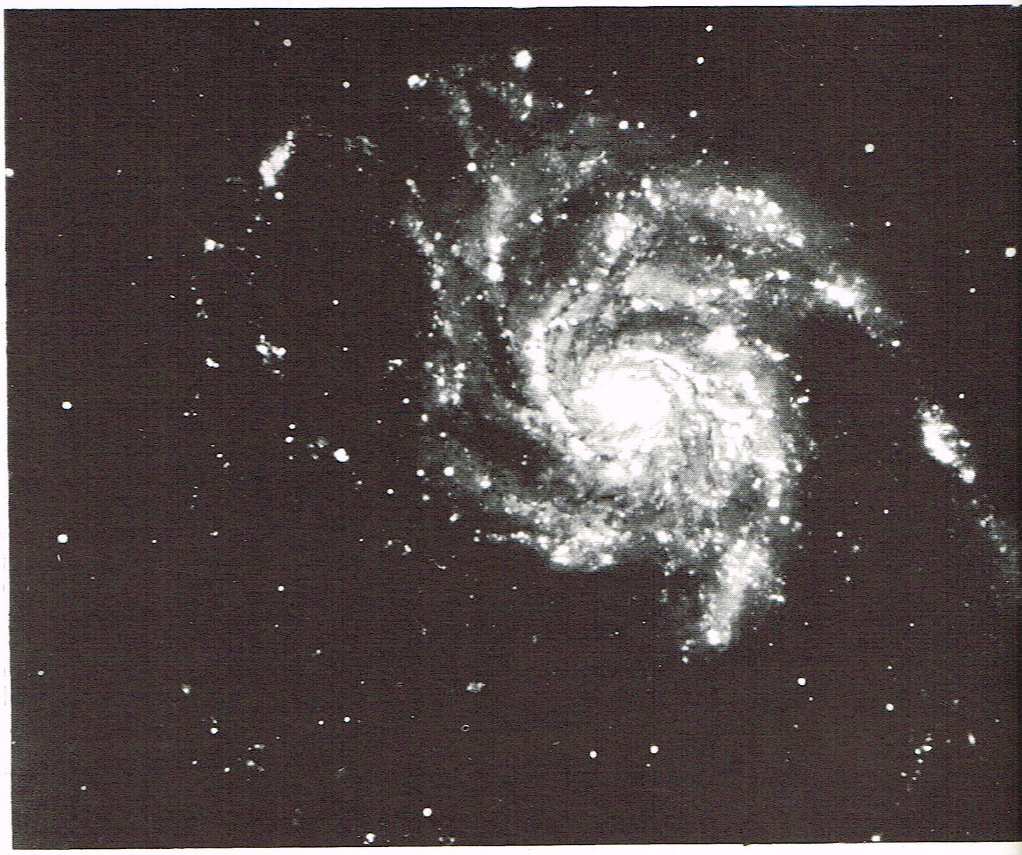


Schéma de la Galaxie (d'après Struve). Le Soleil se trouve à la place indiquée par le petit cercle, à environ 8 000 parsecs du centre de la Galaxie. Il participe à la rotation de celle-ci et se déplace à la vitesse de 250 à 300 km/s, dans la direction indiquée par la flèche.

Ci-dessous, **petite portion d'un cliché pris à l'observatoire du Mont-Wilson. Chacune des taches floues indiquées par une flèche est une galaxie dont les dimensions peuvent être aussi grandes que celles de notre propre Galaxie. Plusieurs centaines de millions, voire des milliards d'années de lumière pour les plus grandes, nous séparent de ces nébuleuses.**











Notre Galaxie peut être également comparée à une roue qui serait renflée en son milieu : en effet, elle tourne sur elle-même. On a évalué son temps de rotation moyen à 200 milliards d'années. N'allons pas croire que notre Soleil en occupe le centre; bien loin de là : il se place modestement dans un de ces prolongements spiralés du bord galactique qu'on appelle des « bras ».

JUSQU'AUX FRONTIÈRES DE L'UNIVERS

par
Anne-Marie ROZIS-SAULGEOT

Avant la construction des grands télescopes actuels, on apercevait les galaxies lointaines sous la forme de nuages brillants, d'où les termes de « nébuleuses » et de « nuages ». Mais le télescope Hale, au Mont-Palomar, dont le diamètre est de 5 m, a permis de « résoudre » ces nébulosités en étoiles distinctes.

Sur les photos prises avec ce télescope géant, on pourrait déceler jusqu'à un milliard de galaxies. Cependant, même avec ce télescope, on constate que les galaxies ne nous présentent généralement que des contours imprécis et qu'il est toujours impossible de voir ou de photographier leurs détails, tant sont grandes leurs distances, évaluées à plusieurs centaines de millions d'années de lumière.

Le ciel nocturne offre à nos yeux, en l'absence de nuages et de clair de lune, un fourmillement d'étoiles grandes et petites. Ainsi nous est révélée l'immensité d'un univers dans lequel brillent des milliards de soleils comparables au nôtre. Bien qu'elles soient souvent plus grandes et plus lumineuses que lui, les étoiles n'ont aucun diamètre apparent et ne se présentent à nous que sous l'aspect de points sans aucun détail perceptible, même dans les plus grands télescopes. Cela tient à l'énorme distance qui les sépare de notre système solaire.

Herschel (1738-1822) fut le premier à réaliser que notre Soleil faisait partie d'une immense famille d'étoiles, ou, si l'on préfère, d'une région de l'univers qu'on appelle la « Galaxie » (notre Galaxie). Au-delà s'étend un espace considérable où ne se trouve aucune étoile. Puis commencent d'autres univers, c'est-à-dire d'autres galaxies. On aurait tort de croire qu'elles sont d'une dimension inférieure à celle de la nôtre. Il n'en est rien, alors même qu'elles apparaissent au bout des plus grands télescopes comme de modestes taches de lumière. Encore une fois, cet amenuisement n'est dû qu'à la distance.

Il y a une façon bien simple de se convaincre de l'importance de notre propre Galaxie : c'est de regarder la Voie lactée, qui en constitue la « tranche » ou la « coupe ». Elle présente, en effet, grossièrement, la forme d'une lentille, d'une montre, d'une galette. Un microbe, placé à l'intérieur d'une galette et qui se dirigerait vers l'un des bords du gâteau, rencontrerait, avant d'en sortir, une grande quantité de matière, alors qu'il traverserait

moins de matière et en sortirait bien plus vite s'il adoptait une direction perpendiculaire. Il en est ainsi de notre regard; lorsque nous le dirigeons dans le plan de la Galaxie, nous voyons un nombre vertigineux d'étoiles, qui, par l'effet de la perspective, nous paraissent serrées les unes contre les autres, au point de ne plus former qu'une nébulosité : c'est la Voie lactée. Lorsque nous regardons dans d'autres directions, nous y voyons bien moins d'étoiles, et elles semblent relativement clairsemées.

Parmi les autres galaxies accessibles à notre observation, nommons la Grande Nébuleuse d'Andromède. Elle est éloignée de nous de 450 000 parsecs (17.10^{18}, c'est-à-dire 17 milliards de milliards de kilomètres), ce qui signifie, comme on le dira plus loin, que sa lumière met 1 500 000 années à nous parvenir, à la vitesse de 300 000 km/s.

Plus proches de nous — à 50 000 parsecs seulement ! — se trouvent les Nuages de Magellan, visibles de notre hémisphère austral.

Contrairement à ce que l'on croyait il y a vingt ans, les galaxies, bien loin d'être solitaires, ont tendance à se trouver groupées en amas. La nôtre n'est qu'un élément du « groupe local ». La grande galaxie de la Nébuleuse d'Andromède (Messier 31) en fait également partie. On connaît une vingtaine d'éléments de ce « groupe local ».

Nous venons d'esquisser un tableau d'ensemble, évidemment très incomplet, de l'univers tel qu'il nous apparaît par une belle nuit sans lune, et tel que l'expliquent les découvertes et les calculs de l'astronomie moderne. Les études se poursuivent. On recherche, par exemple, les astres obscurs ou les planètes qui gravitent autour des étoiles. On a décelé, au centre de notre Galaxie, une très puissante source d'ondes hertziennes dont le mystère n'est pas encore percé. Des travaux sont également en cours sur le « noyau » des galaxies extérieures. Des observatoires se sont créés partout sur notre globe, même dans son hémisphère Sud, car toute une partie du ciel n'est jamais visible sous nos latitudes; c'est là que brille, entre autres constellations « australes », la Croix du Sud.

Au cours de cette introduction, nous avons fait maintes fois allusion à des grandeurs vraiment « astronomiques », et notamment aux distances vertigineuses qui nous séparent des étoiles. Le lecteur est en droit de demander sur quoi sont fondées de telles évaluations, de quelles méthodes elles relèvent et de quel ordre sont les erreurs inévitables. C'est ce sujet qui fera l'objet des pages suivantes.

comment mesurer les distances astronomiques ?

Certaines étoiles nous paraissent fort brillantes, d'autres ont un faible éclat. Cette diversité doit-elle être attribuée à une différence dans leur luminosité réelle ou dans leur éloignement? Aucune réponse immédiate ne peut être donnée à cette question.

Non seulement toutes les étoiles, sauf le Soleil, même observées à travers les télescopes les plus puissants, se présentent à nous comme de simples points lumineux, mais, encore, chacune paraît parfaitement immobile par rapport à l'ensemble des autres (« sphère des fixes »). Ces deux caractères, fixité et aspect ponctuel, expliquent les difficultés considérables que les astronomes de tous les temps ont rencontrées dans l'évaluation des distances stellaires.

Il se trouve cependant qu'un certain nombre d'étoiles ne se contentent pas d'être de simples points lumineux absolument immobiles et défiant toute investigation; certaines d'entre elles nous adressent, semble-t-il, des signaux minimes, difficiles à déceler, mais qui attisent pourtant notre curiosité. Nous allons voir comment ceux-ci nous permettent d'évaluer les distances, tout en nous révélant la nature de l'étoile qui les émet.

les parallaxes trigonométriques, mesures de la « proche banlieue »

Dans un pays de montagne, un ivrogne se promène la nuit dans une rue. A une dizaine de mètres devant lui se trouve un réverbère, qui brille sur le fond sombre constitué par la montagne couverte de sapins. Si notre sympathique soûlard titube, il lui semblera que la lampe se déplace devant le fond apparemment immobile de la montagne. La situation est la même pour les étoiles : le fond montagneux, c'est l'ensemble des astres très lointains, tellement lointains qu'ils nous apparaissent comme absolument immobiles sur la voûte céleste, en dépit des vitesses fantastiques (plusieurs dizaines ou centaines de kilomètres par seconde) qui les animent. L'ampoule du réverbère, c'est une des étoiles relativement proches de nous. Et l'ivrogne qui titube? Eh bien,

c'est notre vieille Terre, qui décrit indéfiniment son va-et-vient annuel tout autour du Soleil, effectuant ainsi un éternel zigzag. Il en résulte que les étoiles relativement proches semblent, elles aussi, osciller de gauche à droite, et *vice versa*, par rapport à la voûte céleste. Et cela se répète régulièrement chaque année. On dit que les étoiles proches décrivent une « ellipse parallactique » sur la sphère céleste.

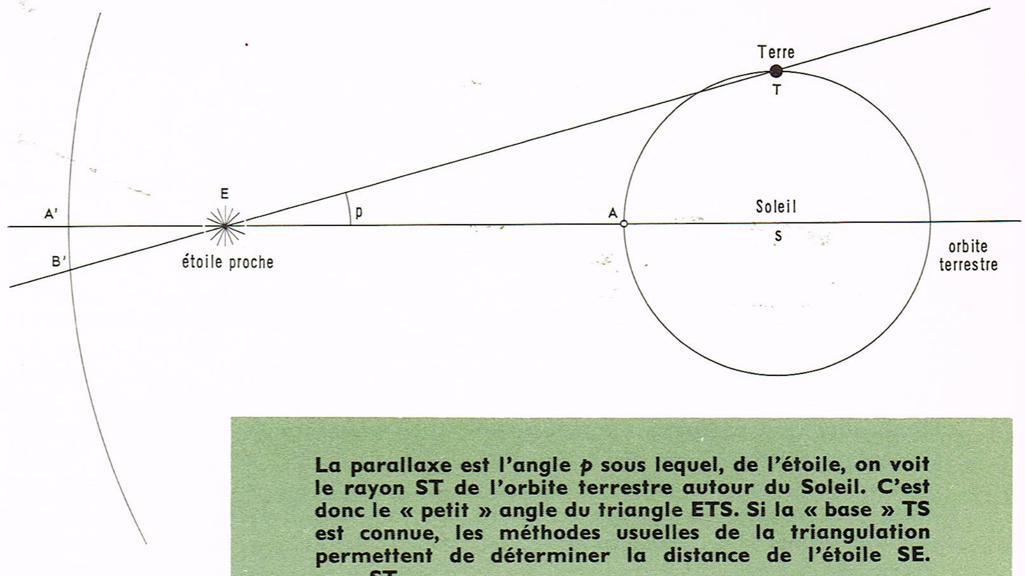
On conçoit facilement les limites d'une telle méthode. Elle ne s'applique pas à la montagne, mais seulement au réverbère; pas au fond du ciel, mais seulement aux étoiles proches. Et même pour ces dernières, les déplacements qu'il faut mesurer pour la mettre en œuvre sont si petits qu'ils paraissent, à première vue, défier le sens commun. Ils sont de l'ordre d'une seconde d'arc (angle sous lequel on verrait un homme à 360 km, ou une pointe d'aiguille à 50 m). Cette faiblesse du déplacement parallactique est due à ce que, le Soleil excepté, les étoiles les plus proches sont à des distances énormes de nous.

Ce n'est qu'en 1838 que Bessel réussit à mettre en évidence le faible déplacement de « 61 Cygni ». Il trouva un tiers de seconde d'arc par an. Il fallut attendre 1900 pour que l'on commence à compter par centaines les étoiles dont le mouvement parallactique était mesurable. On en dénombre aujourd'hui au moins 20 000.

On appelle « parallaxe » l'angle sous lequel on verrait, de l'étoile, le rayon de l'orbite terrestre. Sa connaissance permet de déterminer facilement la distance de l'étoile considérée, du moins si l'on a déterminé une fois pour toutes la distance de la Terre au Soleil.

L'étoile α du Centaure, notre plus proche voisine, a une parallaxe un peu inférieure à une seconde d'arc. Sa distance est de l'ordre de 4.10_{13} km.

La méthode des parallaxes trigonométriques est la seule qui permette une détermination directe de la distance des étoiles. Malheureusement, même les étoiles les plus proches sont à de telles distances, en comparaison de l'orbite terrestre, que le déplacement de l'image de l'étoile est toujours très petit : non seulement les calculs sont longs et fastidieux, mais encore, dans ces conditions, les erreurs sont toujours du même ordre de grandeur que les résultats.



La parallaxe est l'angle p sous lequel, de l'étoile, on voit le rayon ST de l'orbite terrestre autour du Soleil. C'est donc le « petit » angle du triangle ETS. Si la « base » TS est connue, les méthodes usuelles de la triangulation permettent de déterminer la distance de l'étoile SE. $d \sim \frac{ST}{p}$; ST = 150.10^6 km; pour $p = 1''$ d'arc, on trouve $d = 1$ parsec = $3{,}1.10^{13}$ km.

Il n'est donc pas étonnant que les programmes de recherches de nouvelles parallaxes soient pratiquement abandonnés : on ne peut espérer autre chose que des indications approchées sur quelques milliers d'étoiles relativement voisines. Mais ces parallaxes trigonométriques, si imprécises soient-elles, servent à étalonner tous les autres indicateurs de distance.

La « base » employée en astronomie est le rayon de l'orbite presque circulaire que la Terre décrit autour du Soleil. On conçoit dès lors que cette constante ait une importance fondamentale; pour cette raison on l'appelle « unité astronomique ». Sa détermination est un problème essentiel. Elle n'est pas possible directement, mais seulement à partir de l'observation de petites planètes telles qu'Erôs. Sa valeur est voisine de 150 millions de kilomètres. (Récemment, on a pu envoyer un « écho-radar » sur Vénus et déduire de la durée du trajet aller et retour une excellente valeur de l'unité astronomique.)

L'unité astronomique de distance se révèle, en fait, beaucoup trop petite pour être utilisable. La distance de l'étoile la plus proche, α du Centaure, est déjà deux cent mille fois supérieure. Aussi emploie-t-on une unité beaucoup plus grande, dont la définition est étroitement liée à celle de la parallaxe, le *parsec*. C'est la distance à laquelle le rayon de l'orbite terrestre apparaît sous un angle de une seconde d'arc. Il revient au même de dire qu'une étoile placée à 1 parsec a une parallaxe de 1 seconde. Le parsec vaut $3{,}1 . 10^{13}$ km (environ deux cent mille unités astronomiques). Ainsi, la distance de l'étoile α du Centaure est légèrement supérieure à 1 parsec. Tout comme l'unité astronomique, le parsec n'est bien défini que dans la mesure où le rayon de l'orbite terrestre autour du Soleil est parfaitement connu. Il n'en est malheureusement pas ainsi, et l'on peut dire que l'évaluation des distances des étoiles en parsecs donne essentiellement des valeurs relatives.

Pour s'affranchir de cet inconvénient, on a cherché une unité de distance directement liée à un phénomène physique mesurable; ainsi est née l'*année de lumière*, qui est la distance parcourue par la lumière dans le vide pendant un an. La vitesse de la lumière, dans le vide, est de 300 000 km/s. En un an, celle-ci parcourt une distance voisine de 10^{13} km. *Le parsec vaut sensiblement 3,1 années de lumière*. L'une et l'autre unité sont fort bien adaptées à la mesure des distances dans notre Galaxie. Mais, au-delà, nous trouvons des dimensions d'une telle envergure qu'elles se comptent en millions et en milliards d'années de lumière. On n'a pas encore créé de nouvelle unité de distance adaptée à ces régions très lointaines.

Les limites de la méthode des parallaxes trigonométriques se situent aux environs de 100 parsecs. Comment aller au-delà? Citons tout d'abord une méthode qui n'est qu'une extension de la précédente. On utilise une « base » plus grande que la distance Terre-Soleil : par exemple, l'arc décrit par le Soleil dans sa rotation autour

du centre de la Galaxie pendant quelques dizaines d'années. Cela présente toutefois un petit inconvénient : pendant une si longue période, toutes les étoiles se déplacent de façons diverses; on ne peut espérer par un tel procédé que les résultats statistiques fondés sur la possibilité de négliger les mouvements mutuels, qui s'annulent mutuellement en vertu de la loi des grandes séries.

Toutes les autres mesures de distances, en particulier celles qui nous permettent d'aller bien au-delà des 100 parsecs de la méthode trigonométrique, nécessitent qu'une étoile nous adresse quelque signal particulier, nous permettant de la reconnaître et de la comparer à des étoiles semblables, mais plus proches, dont la distance aura été évaluée par la méthode des parallaxes trigonométriques.

de plus en plus loin dans la Galaxie

Avant d'étudier la nature de ces signaux, il nous faut examiner une question essentielle : l'influence de la distance sur la luminosité apparente des étoiles.

Deux étoiles rigoureusement identiques, mais placées à des distances différentes, ne nous paraissent pas également brillantes. Sans entrer dans les détails qui relèvent de la technique photométrique, disons simplement qu'on étend aux mesures stellaires une loi fondamentale qui a toujours été vérifiée sur la Terre : l'éclat apparent d'une étoile varie comme l'inverse du carré de la distance à laquelle elle se trouve.

Les étoiles constituent un échantillonnage infiniment varié de points lumineux parmi lesquels aucun repère ni aucune graduation ne s'imposent. C'est un peu comme le son d'une sirène qui tourne de plus en plus vite; il paraît de plus en plus aigu, mais il n'existe pas d'échelons naturels de hauteur sonore. Pour en créer artificiellement, on a procédé pour l'éclat des étoiles comme pour les sons : on a pris pour unité la plus petite différence que nos sens puissent percevoir. Ainsi sont nées les diverses « grandeurs » d'étoiles, qui sont à l'origine de l'unité actuellement employée : la *magnitude*. Il nous suffira de savoir que si deux étoiles diffèrent de cinq magnitudes l'une d'elles est cent fois plus brillante que l'autre. Une fois fixé l'échelon de mesure, il nous faut encore préciser l'emplacement de notre « échelle ». Eh bien, le zéro de l'échelle des magnitudes est arbitraire. A l'origine, il avait été conventionnellement fixé que l'étoile polaire aurait la magnitude + 2. On s'aperçut ensuite que la luminosité de cet astre présentait de légères variations; c'est pourquoi on préfère actuellement utiliser à cet effet, pour la détermination de cette grandeur, non plus une étoile unique, mais toute une série d'étoiles de comparaison bien étudiées (séquence polaire).

Les quelques exemples que nous allons maintenant donner permettront au lecteur de se familiariser avec un caractère un peu déroutant de ce système : les astres les plus brillants ont les magnitudes les plus faibles, et ces valeurs peuvent même être négatives dans les cas extrêmes; c'est ainsi que la magnitude de Sirius, l'étoile la plus brillante du firmament, vaut — 1,50; celle de α du Centaure, moins brillante, à peu près 0; celle de la plupart des étoiles qui constellent le ciel nocturne, entre 2 et 6. Au-delà, à partir de 7, les corps célestes ne sont plus visibles à l'œil nu. Il est possible d'attribuer une magnitude apparente au Soleil et à la pleine Lune : elle est respectivement de — 26 et — 14. A l'opposé, la limite actuelle des étoiles les plus faibles que l'on puisse déceler est de l'ordre de 20 magnitudes, ce qui correspond à un éclat cent millions de fois plus faible que celui de l'étoile de magnitude zéro. Cette limite n'est imposée que par la luminosité du ciel nocturne, sur lequel ne se détachent pas, même par les nuits les plus noires, les objets trop peu lumineux. **C'est précisément pour se débarrasser de cette limitation que l'on envisage de monter télescopes et instruments à bord de fusées ou de satellites spécialement conçus (astronomie extra-terrestre).**

Si nous voulons déterminer la luminosité intrinsèque, il nous faut éliminer le facteur « distance ». On convient, pour cela, de calculer la magnitude qu'aurait l'étoile si elle était située à 10 parsecs de la Terre. On appelle « magnitude absolue » la valeur ainsi calculée. Cette notion est fondamentale en astrophysique. Sa détermination permet de calculer le flux global de lumière déversé par l'étoile dans l'espace, d'en déduire quels sont sa température de surface et son diamètre, et de fixer un grand nombre de ses propriétés physiques.

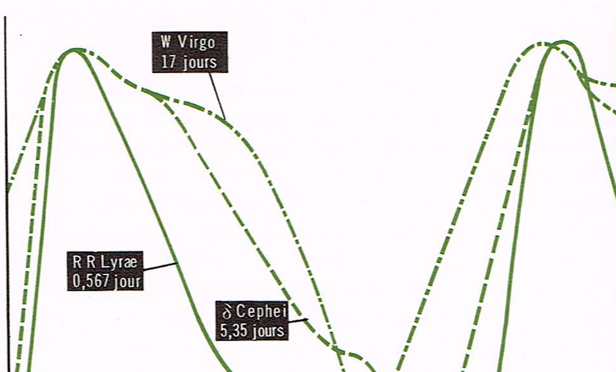
Un mot maintenant sur les méthodes qui nous permettent d'aller au-delà de cette limite de 100 parsecs imposée à la méthode trigonométrique. Les procédés sont divers, mais ils utilisent tous le même principe général. **Si l'on admet que la magnitude absolue d'une étoile est fixée d'une manière unique par ses particularités spectrales, on peut, connaissant ces dernières, en déduire cette magnitude absolue et la distance entre l'étoile et nous.**

La méthode nécessite des études extrêmement laborieuses. Elle n'est applicable qu'aux objets suffisamment rapprochés pour qu'une analyse spectrographique puisse être menée à bien; aussi l'appelle-t-on « méthode des parallaxes spectroscopiques ».

des proches galaxies aux frontières de l'Univers

Nous trouvons ici un cas typique de ce que nous avons appelé les signaux que nous font certaines étoiles comme pour nous intriguer.

En effet, un assez grand nombre de ces astres ont une propriété curieuse; leur luminosité varie régulièrement, la période constante de cette variation pouvant aller de un à quelques jours. De telles oscil-



Courbes de lumière de plusieurs étoiles pulsantes : RR Lyrae, δ Cephei et W Virgo.

lations sont liées à une structure particulière, caractérisée par une pulsation radiale vers l'intérieur et vers l'extérieur.

Deux groupes d'étoiles variables particulièrement caractéristiques sont utilisés comme indicateurs de distance. Le premier groupe porte le nom de leur prototype (*RR de la Lyre*), bien que l'on connaisse des milliers d'étoiles de ce type. Cependant, en dépit de leur dispersion dans toutes les régions du ciel, ce groupe d'étoiles est physiquement très homogène : luminosité cent fois supérieure à celle du Soleil, période courte, de l'ordre d'un jour. Elles permettent de mesurer des distances atteignant 200 000 parsecs.

Le deuxième groupe d'étoiles variables, très intéressant, est celui des *céphéides*, du nom de l'étoile δ Céphée. Extrêmement lumineuses et facilement reconnaissables, elles forment un groupe beaucoup moins homogène. Leurs périodes sont diverses, allant de 60 heures à 35 jours. En faisant l'hypothèse qu'à chaque période correspond une structure particulière, on admet l'identité de toutes les céphéides ayant la même période. On dispose ainsi d'un précieux indicateur de distance.

Les étoiles RR de la Lyre et les céphéides sont utilisées pour faire des mesures non seulement à l'intérieur de notre Galaxie, mais encore en dehors d'elle. C'est ainsi qu'on a pu préciser les indications que nous avons données concernant la Grande Nébuleuse d'Andromède et les Nuages de Magellan.

Admettant aussi que les galaxies de même structure ont le même éclat, une méthode analogue permet d'estimer les plus lointaines distances, jusqu'au milliard d'années de lumière des nébuleuses que découvre le télescope géant du Mont-Palomar. **On atteint ici même la frontière de l'univers optiquement connu.**

de la détermination des distances à la compréhension de l'Univers

La détermination des distances et des magnitudes absolues est, à n'en pas douter, un des problèmes fondamentaux de l'astronomie. Elle permet seule d'aborder des questions d'une vaste envergure : esquisser la structure de la Galaxie, en identifier les bras spiraux, évaluer la densité générale de l'Univers (constante essentielle en cosmologie).

Les problèmes concernant l'évolution des étoiles et leurs vitesses intrinsèques sont également liés à la détermination des distances et des magnitudes absolues.

Donnons quelques exemples : on a établi que le centre de notre Galaxie est distant de 8 000 parsecs. (Il serait peut-être plus suggestif de dire que la lumière met 25 000 ans pour franchir cette distance.)

L'étude des étoiles variables nous permet également d'évaluer la distance à laquelle se trouvent d'autres galaxies. Sandage estime à 2 500 000 parsecs la distance de l'amas de la Vierge. Les limites actuelles atteintes par cette méthode sont de l'ordre de 10 millions de parsecs. Les distances « ultimes » qui intéressent les astronomes se comptent par milliards de parsecs (pour nous parvenir de ces galaxies très lointaines, la lumière met plusieurs milliards d'années, temps comparable à l'âge qu'on attribue au système solaire).

D'immenses éloignements, un nombre considérable d'astres groupés en galaxies et en amas, et, entre ceux-ci, des vides encore plus grands; tel est le tableau que l'astronomie du XXe siècle offre à nos regards.

Ajoutons que le tout est entraîné dans un tourbillonnement rapide et complexe.

Pour achever ce tableau, disons que les galaxies lointaines semblent nous fuir à une allure folle, de sorte que le monde ne cesse de se dilater. C'est ce qu'essaient d'expliquer des théories comme celle de l'« expansion de l'Univers », que nous examinerons en leur temps.

les performances des grands télescopes

Après avoir parlé des distances et des problèmes que pose leur mesure, il est bon de dire quelques mots des instruments qui permettent de mener à bien l'exploration astronomique de l'Univers. Ce sont essentiellement les télescopes, dont la construction et la mise au point demandent des années, qui ont apporté la plus large contribution aux progrès de la science. Il faut mentionner aussi les lunettes astronomiques, dont la fin du XIXe siècle connut l'heure de gloire, mais dont la construction est pratiquement abandonnée de nos jours.

La différence essentielle entre ces deux types d'instruments résulte du fait que l'objectif de la lunette est constitué par des lentilles accolées, alors que la partie principale d'un télescope est un large miroir concave. Plus l'instrument est de grande surface, plus il récolte de lumière; ces appareils sont en effet de véritables « entonnoirs » à lumière (c'est donc essentiellement leur diamètre — la surface de leur miroir — qui détermine leurs performances). Il est malheureusement impossible, pour des raisons techniques, de construire des lunettes dont l'objectif ait un diamètre supérieur à un mètre (observatoire Yerkes). Nous allons voir que les télescopes permettent en revanche d'atteindre des dimensions beaucoup plus grandes, ce qui a valu à ces derniers une victoire définitive dans la course aux performances. (La situation géographique d'un instrument exerce elle aussi une grande influence sur les résultats que l'on peut en attendre; nous reviendrons, en fin de chapitre, sur ce point important.)

La partie essentielle d'un télescope est un grand miroir concave, qui fait converger la lumière provenant d'une étoile en un point appelé « foyer », où se forme l'image de l'étoile. Tous ces miroirs sont généralement en verre, leur surface ayant été rendue réfléchissante par un dépôt métallique.

Parmi les parties accessoires de l'instrument, citons d'abord le « tube », qui n'a d'autre utilité que d'éviter la lumière parasite; dans la plupart des télescopes, un second miroir, beaucoup plus petit, renvoie l'image définitive sur le côté du tube, où l'on place la « cage » de l'opérateur. (Le télescope Hale du Mont-Palomar a un diamètre si grand que ces dispositifs ne sont plus nécessaires : la cage de l'opérateur, qui n'arrête plus

qu'une faible fraction de la lumière incidente, est tout simplement placée au milieu du tube où se trouve le foyer.) Voici quels sont les principaux avantages qui ont donné la victoire au télescope sur la lunette :

1° Facilité relative de polir une grande surface ;

2° Possibilité d'utiliser un verre dont les propriétés optiques ne sont pas parfaitement homogènes dans toute la masse, ce qui est absolument exclu pour les objectifs de lunettes, que la lumière traverse au lieu de s'y réfléchir seulement ;

3° Absence de toute aberration au foyer, du moins pour un télescope parabolique;

4° Possibilité de soutenir le miroir par toute sa surface inférieure, tandis qu'une lentille ne peut être soutenue que par ses parties périphériques.

On peut dire que la première moitié du XXe siècle a vu une véritable course aux télescopes géants.

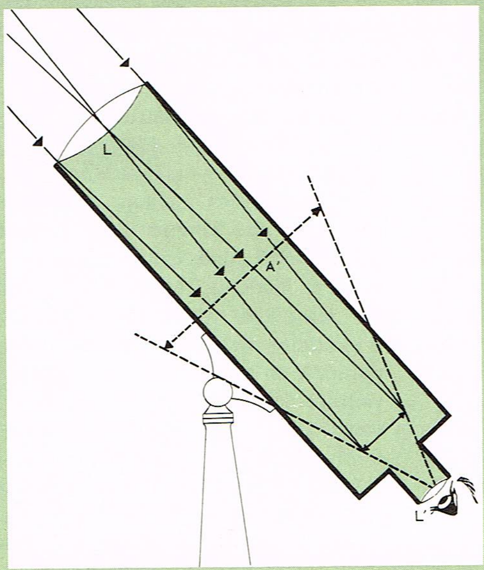
A tout seigneur tout honneur! Nous citerons donc tout d'abord le *télescope Hale du Mont-Palomar* (en fonctionnement depuis 1948), situé en Californie, à une altitude de 1 700 m. Son diamètre dépasse légèrement les 5 m. Pour donner une idée de ce mastodonte, indiquons que le tube du télescope pèse 140 tonnes. Le miroir est fait en Pyrex. Un bloc de verre de 20 tonnes a dû être coulé pour qu'on pût tailler ce miroir, et il n'a pas fallu moins de sept ans pour en assurer le complet refroidissement!

C'est grâce à lui qu'il a été possible de « résoudre » les nébuleuses lointaines et de mener à bon terme la plupart des travaux récents sur les galaxies éloignées, dont nous avons dit un mot au chapitre précédent et dont nous reparlerons à propos de l'évolution de l'Univers.

Après ce géant, nous devons citer le *télescope du Mont-Wilson*, dont le diamètre est de 2,50 m. Lui aussi se trouve en Californie, à une altitude sensiblement égale à celle du précédent. Si le géant du Mont-Palomar mérite la première place par son diamètre, le télescope du Mont-Wilson mérite peut-être d'être classé *ex aequo* avec lui, car, étant en service depuis 1917, il a contribué puissamment aux progrès de l'astronomie pendant la première moitié de ce siècle.

Depuis peu (1959), il existe, toujours aux Etats-Unis, un grand télescope de 3 m : c'est celui de l'observatoire Lick.



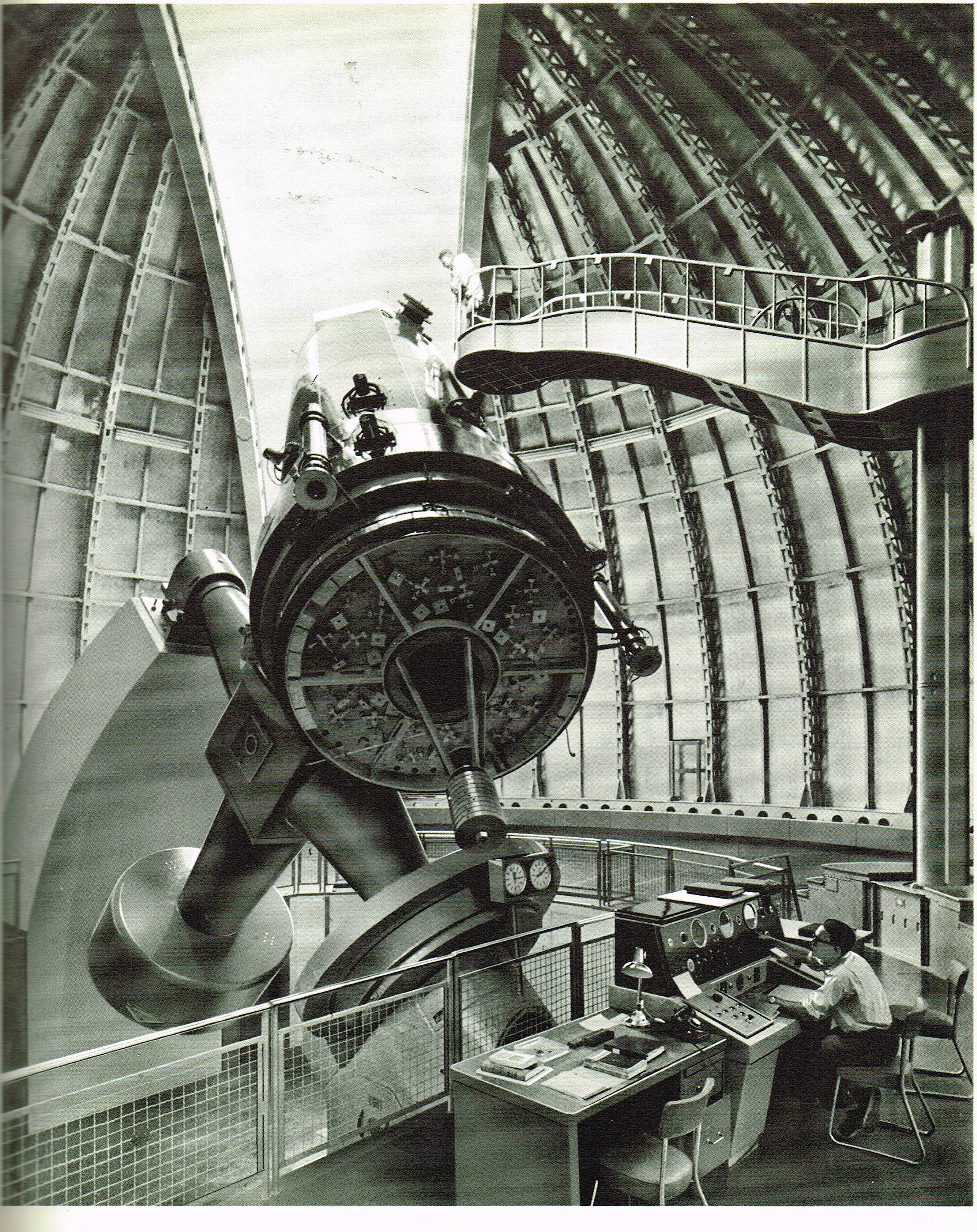


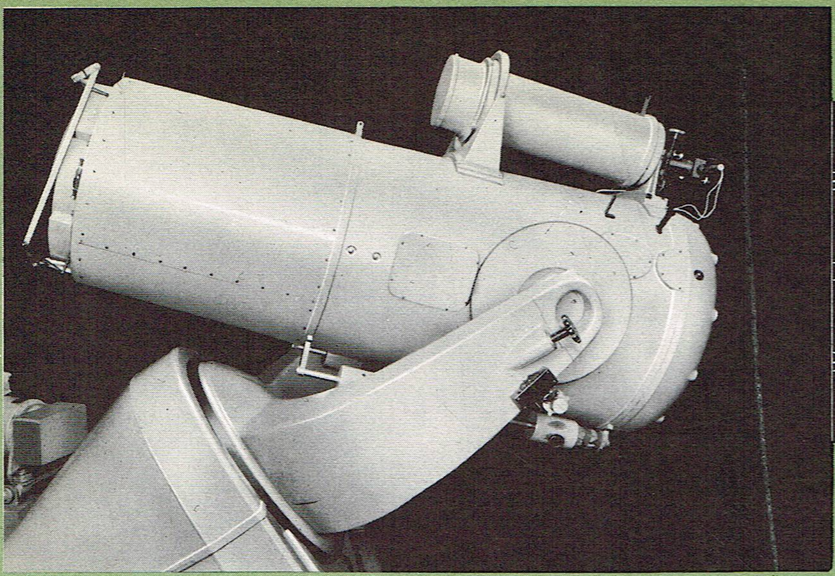
La grande lunette de l'observatoire Yerkes (États-Unis), longue de 19 m, qui, avec un mètre d'ouverture, est la plus grande jamais construite.

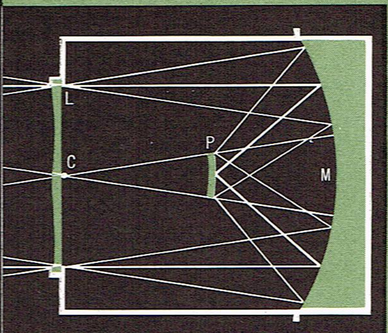
Principe de la lunette astronomique. L'objectif L donne d'un objet à l'infini une image réelle renversée, dont l'oculaire donne une image virtuelle, vue agrandie en A'.

***Page ci-contre,* le télescope de 193 cm de l'observatoire de Haute-Provence.**

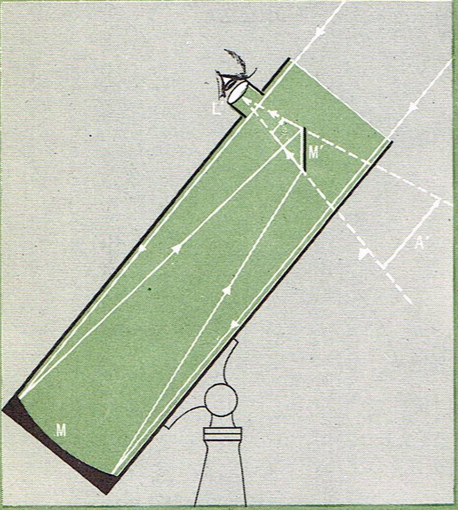
Phot. Ellerman, O. H. P.



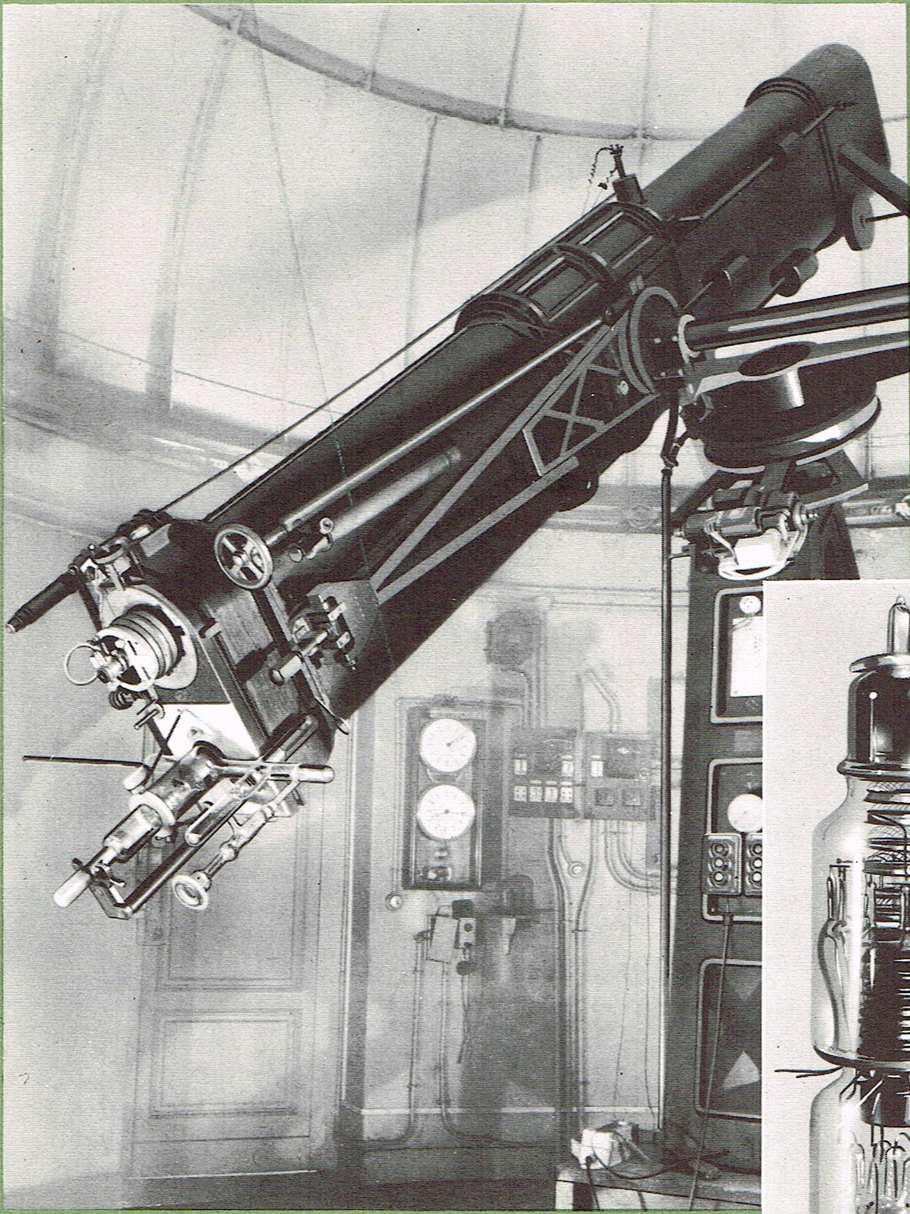


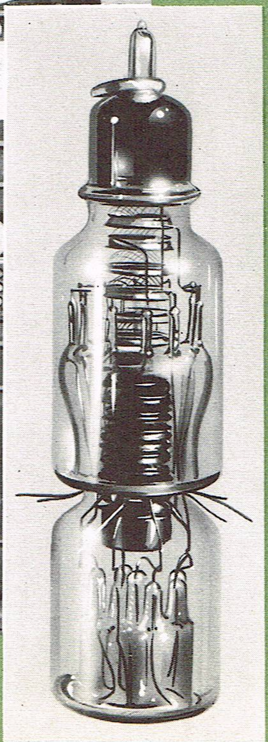


La chambre de Schmidt est un télescope spécialement conçu pour la photographie pour lequel on se sert d'un miroir sphérique M, beaucoup plus facile à réaliser que le miroir parabolique. L'aberration de sphéricité est corrigée par la lentille asphérique L placée au centre de courbure C. Une image nette, d'un champ très étendu, se forme sur la surface sphérique du porte-film P. *A gauche*, la chambre de Schmidt du Mont-Palomar (0,45 m d'ouverture).



Principe du télescope. Le miroir parabolique M donne d'un objet à l'infini, après réflexion sur le petit miroir plan M', une image réelle, dont l'oculaire L' fournit une image virtuelle, vue agrandie en A'.





La caméra électronique de Lallemand.

Le principe de cette caméra électronique, que l'on appelle improprement « télescope », est comparable à celui de la télévision. En fait, ce n'est pas un télescope, mais un accessoire que l'on fixe au bout d'un télescope ordinaire afin de recueillir et d'amplifier la lumière « par voie électronique ». On peut également fixer la caméra à l'extrémité d'un télescope équipé d'un spectrographe; elle amplifie alors la luminosité du spectre, son image est reçue sur une plaque sensible, sorte de cellule photo-électrique, qui émet des électrons en chaque point où la lumière la frappe. Ces électrons, judicieusement canalisés et multipliés, sont recueillis sur une plaque photographique spéciale, où ils forment l'image du spectre. Le principe de cet appareil relève de l'« optique électronique ». Il est comparable à celui de la télévision ou du microscope électronique.

La France, qui occupe une place honorable avec le télescope de 193 cm de l'observatoire de Haute-Provence, construit actuellement un instrument de 350 cm.

La situation géographique des observatoires où sont installés les télescopes détermine en grande partie les performances de ces instruments. C'est ainsi que les savants français, avec une modeste lunette de 60 cm de diamètre placée d'une façon tout à fait remarquable au sommet du pic du Midi de Bigorre, ont pu obtenir certains résultats « compétitifs » avec ceux des Etats-Unis et de l'U. R. S. S. — notamment sur la structure des planètes. En effet, les conditions climatiques jouent un rôle fondamental. Il faut éviter, en particulier, la scintillation due à la turbulence atmosphérique; la netteté des images est aussi fortement compromise par l'humidité de l'air et par les courants de convection susceptibles de se produire autour de la coupole protégeant le télescope. La pollution atmosphérique, le brouillard sont autant d'obstacles à une bonne observation astronomique. Enfin, bien entendu, les facteurs climatiques proprement dits sont essentiels, le travail astronomique ne se faisant jamais que par beau temps.

Enfin, la question de la latitude a une grande importance : alors qu'un télescope situé au pôle Nord ne permettrait de voir qu'une petite partie de l'hémisphère correspondant, un instrument situé à l'équateur ne peut étudier qu'une partie des étoiles des deux hémisphères; enfin, le groupement de Magellan n'est observable qu'à partir d'instruments situés suffisamment bas dans l'hémisphère austral. Il existe déjà de tels observatoires (Bloemfontein, Good Hope, Pretoria en Afrique; Canberra en Australie), mais en nombre encore insuffisant. L'Union astronomique internationale a patronné la création de deux grands observatoires au Chili.

Ainsi le globe terrestre est parsemé d'observatoires juchés le plus souvent sur des sommets, et l'on peut dire qu'une pacifique armée de télescopes est braquée en permanence sur le ciel pour y percevoir les moindres événements. Les astronomes, ces vigies de l'espace, doivent faire preuve, à notre époque, d'endurance, de patience et d'un inlassable dévouement à la recherche de la vérité. Leur existence comporte de perpétuelles veilles nocturnes, dans des coupoles non chauffées, même en hiver, malgré la température généralement très basse des hautes altitudes (le chauffage engendrerait des courants aériens qui troubleraient les observations).

Après leur avoir rendu ce bref hommage, souvenons-nous aussi de la cohorte des inventeurs, opticiens, constructeurs souvent artisanaux, dont le labeur et l'ingéniosité n'ont cessé, depuis quelques siècles, d'améliorer les instruments matériels de la recherche. Sans eux, point d'observations, et si belles que soient les théories, elles ne peuvent s'édifier valablement que sur des faits dûment constatés.

Mais là ne s'arrête pas le travail des astronomes, car, après l'observation et la prise des clichés photographiques, ils doivent se consacrer à l'étude de documents qui peuvent comporter chacun des dizaines de milliers d'informations. Tâche longue et fastidieuse que d'interpréter un seul cliché, et des semaines de travail n'y suffisent pas toujours. Tâche encore plus longue lorsqu'il s'agit des milliers de clichés de la carte du ciel ou des dénombrements d'objets extra-galactiques. Mais c'est à ce prix que sont obtenues toutes les données dont l'étude systématique nous permet de connaître la position, la vitesse, la structure des étoiles, leur état physique et chimique, leurs différences et leurs analogies.

Le télescope n'est pas l'arme unique de l'astronome dans sa chasse aux images. De nombreux appareils viennent le compléter, permettant une moisson encore plus fructueuse. Nous parlerons longuement dans les prochains chapitres du spectrographe; il serait injuste de ne pas citer aussi le coronographe, auquel nous devons la plupart de nos connaissances sur la surface du Soleil, la caméra électronique, le microscope à papillotement pour l'étude des étoiles variables; il faut surtout citer l'énorme développement de la radio-astronomie, qui, avec ses radiotélescopes géants, nous ouvre la porte d'un monde inconnu.

le spectre, message essentiel

Les corps célestes ne sont pas des objets de science analogues aux autres : **l'homme est incapable de les modifier pour mieux les étudier. La science des astres est, et reste, une science de pure observation.**

Pendant des millénaires, l'astronomie s'est bornée à examiner les positions des corps célestes et leur cheminement sur les routes du ciel. C'est seulement au XIXe siècle qu'elle s'est occupée de leur rayonnement.

On peut classer les objets brillants que l'astronomie découvre dans le ciel en trois groupes :

1° Ceux qui sont lumineux par eux-mêmes : Soleil et étoiles;

2° Les planètes, qui réfléchissent la lumière du Soleil;

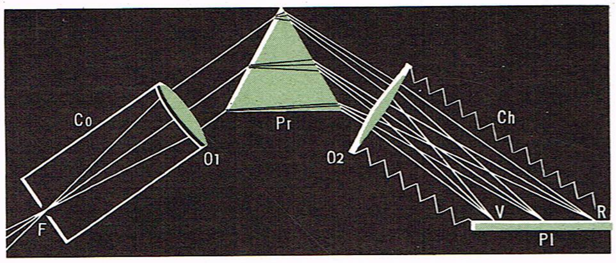
3° Les nébuleuses, dont la luminosité semble due à une sorte de « fluorescence » provoquée par une émission d'ondes ultraviolettes dues à un astre voisin.

La lumière qu'émettent les corps célestes n'est pas simple. Depuis la célèbre expérience de Newton (1666), on sait décomposer à l'aide d'un prisme toute lumière complexe en ses constituants, c'est-à-dire en radiations simples et monochromatiques (v. au tome II, chap. *La physique des ondes*). L'image qui résulte de l'étalement, sur un écran, de ces diverses lumières colorées, s'appelle un « spectre ». Dans le cas de la lumière solaire, on voit s'échelonner sept couleurs allant du rouge au violet. De plus, on détecte, des deux côtés de ce spectre, ces radiations invisibles que sont l'infrarouge et l'ultraviolet.

quelques spectres caractéristiques

L'analyse spectrale peut se pratiquer sur toute sorte de lumière. On trouve alors, selon les différentes sources d'éclairement, des spectres très différents. On a pu établir qu'à chaque élément chimique porté à l'incandescence correspondent des séries de « raies » caractéristiques.

Il est donc possible d'établir avec une quasi-certitude la présence ou l'absence d'un élément dans un astre donné. La spectrographie appliquée aux corps célestes permet en outre d'étudier plusieurs facteurs d'ordre physique les concernant. De tous ces sujets, l'homme n'avait auparavant aucun moyen d'investigation. **Il**



Principe du spectrographe à fente. La lumière concentrée par le télescope sur la fente F du collimateur Co et dispersée par le prisme Pr est recueillie dans la chambre Ch sur la plaque Pl où se forme l'image très étalée du spectre du violet V au rouge R. L'objectif O_1 du collimateur qui rend le faisceau parallèle est achromatique; l'objectif O_2 de la chambre peut être une lentille simple, mais, dans ce cas, la plaque doit être placée obliquement pour compenser l'aberration chromatique de la lentille.

n'est donc pas exagéré de dire que le spectre est le message essentiel que nous envoie le Cosmos. Kirchhoff (1824-1887) formula de grandes règles qui n'ont jamais été mises en défaut :

A) *Les solides incandescents émettent des spectres continus, c'est-à-dire des spectres où la lumière est régulièrement répartie entre les diverses couleurs ;*

B) *Les gaz incandescents, sous basse pression, donnent des lignes brillantes ;*

C) *Un gaz placé devant une source émet un spectre continu barré de raies noires. On dit que ce sont des raies d'absorption. En effet, la lumière émise par la source est absorbée par le gaz, mais seulement dans certaines régions précises du spectre — d'où les raies obscures.*

Une nébuleuse brillante étant formée d'une substance gazeuse, son spectre sera fait de lignes brillantes (cas B). On admet que le centre d'une étoile émet à la façon d'un noyau incandescent (cas A), tandis que les couches gazeuses périphériques, moins chaudes, jouent le rôle d'absorbant pour certaines couleurs déterminées; ce qui provoque des raies obscures (cas C).

C'est également au XIXe siècle que les astrophysiciens ont su reconnaître la nature chimique des substances contenues dans les corps célestes. Ils y ont trouvé partout des éléments qui étaient familiers à notre monde terrestre : hydrogène, fer, nickel, calcium. Les éléments qui, sur notre globe, sont dits « rares », manquent généralement dans les atmosphères stellaires. Ainsi, dès le début des études spectrographiques, les hommes furent mis en face d'une notion grandiose, souvent pressentie auparavant : **l'univers est chimiquement « un ». La matière universelle est partout la même.** Ce fut là une des découvertes les plus importantes de la science moderne.

Cependant, il ne fut pas toujours facile d'identifier les éléments stellaires responsables des raies spectrales : on eut des surprises. Par exemple, on découvrit sur le Soleil un gaz jusqu'alors inconnu, qu'on baptisa « hélium »; plus tard, on trouva que ce gaz existait dans l'air que nous respirons. Dans les nébuleuses, on détecta de fortes radiations dont on ne connaissait pas l'équivalent sur la Terre. On appela « nébulium » le gaz encore inconnu qui les engendrait, ... et puis, on reconnut qu'il s'agissait d'éléments usuels, tels que l'azote et le fer, émettant dans des conditions physiques spéciales (ionisation) qui ne sont jamais réalisées pour eux sur notre globe.

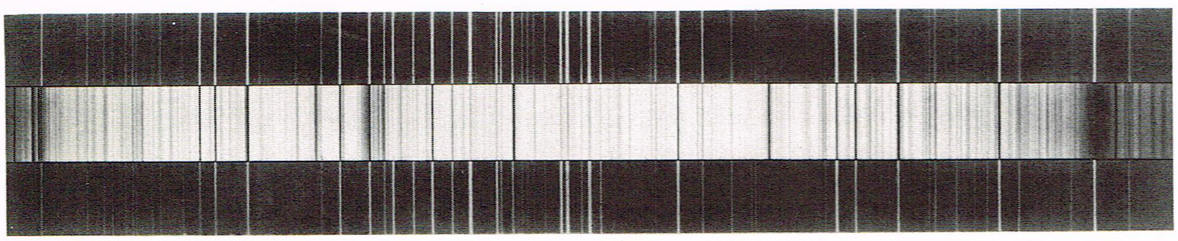
Hâtons-nous d'ajouter qu'en dépit de nombreuses difficultés la carte d'identité spectrale de tous les éléments terrestres est actuellement établie. Leur présence ou leur absence dans les corps célestes proches ou lointains continue à faire l'objet d'innombrables travaux.

spectrographes et réseaux

Les appareils destinés à enregistrer des spectres sont désignés sous le nom de *spectrographes.*

De nos jours encore, on utilise le dispositif de Newton à peu près inchangé dans les principaux domaines de la spectrographie. Toutefois, le prisme de quartz est préféré à tous les autres, parce qu'il n'absorbe pas les radiations invisibles de l'infrarouge et de l'ultraviolet.

Dans les spectrographes modernes, le prisme est souvent remplacé par un dispositif assez différent, le *réseau.* On appelle ainsi une plaque transparente dont la surface est finement striée de lignes parallèles très serrées. La lumière qui traverse un réseau se décompose, à peu près comme à travers un prisme. Jusqu'à ces dernières décennies, les astrophysiciens ont toujours utilisé des prismes, de préférence aux réseaux, parce que ces derniers donnent des images moins lumineuses. Mais, depuis quelques années la fabrication des réseaux a fait de grands progrès. On a obtenu de larges surfaces très finement striées (1 000 traits au millimètre). Comme il arrive toujours, ce progrès instrumental a ouvert à la science de vastes perspectives. On peut désormais découvrir dans les spectres des détails infiniment ténus. De plus, comme nous le verrons plus loin,



Spectre solaire (au milieu) posé sur le spectre du fer (fond sombre). La coïncidence des raies d'émission du fer avec autant de raies d'absorption figurant sur le spectre de la lumière solaire témoigne de l'existence de la vapeur de fer dans la couche renversante du Soleil.

on sait maintenant amplifier dans de grandes proportions la luminosité des images fournies par les réseaux.

Le grand télescope de Saint-Michel en Provence, qui mesure 1,93 m de diamètre — l'un des plus grands d'Europe —, est équipé d'un réseau très puissant. Il permet aux astronomes français d'effectuer des travaux comparables à ceux des astronomes américains.

Une fois le spectre obtenu, il faut l'étudier. La première opération consiste à le photographier. Nous avons donné ci-dessus la reproduction du spectre solaire et du spectre du fer avec leurs raies caractéristiques. L'avantage de la photographie saute aux yeux : ce phénomène fragile qu'est le spectre se trouve ainsi fixé. On peut le conserver indéfiniment, y faire des mesures précises, le comparer à d'autres spectres, etc.

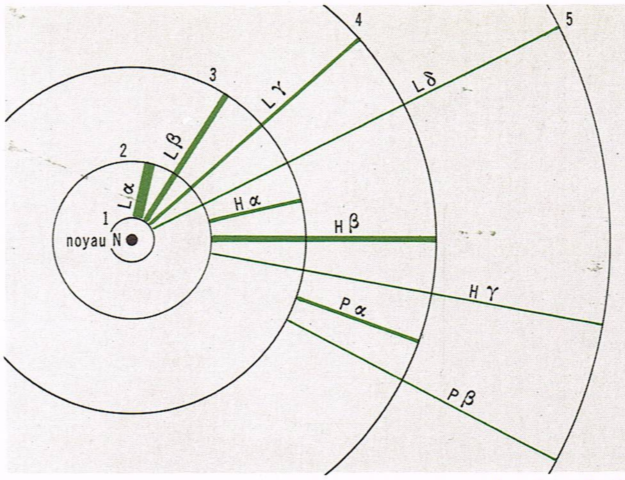
et maintenant, un peu de théorie

Les raies spectrales sont caractéristiques de la lumière émise ou absorbée par une espèce donnée d'atomes. L'emplacement des raies dans le spectre pose un problème complexe, qui est lié à celui de la structure atomique.

Prenons un exemple : les *raies de Fraunhofer du Soleil.* On sait maintenant qu'il s'agit de certaines raies du fer, et surtout de l'hydrogène. En effet, l'hydrogène, à basse pression et soumis à une forte décharge électrique, présente des lignes brillantes qui occupent dans le spectre la même place que les raies noires du spectre solaire. On distingue tout d'abord quatre raies dans la partie visible du spectre (une dans le rouge, une dans le bleu, deux dans le violet). Puis, lorsqu'on étudie ce spectre sur une plaque photographique, on y découvre, au-delà du violet, toute une série de raies situées dans l'ultraviolet, et qui vont en se resserrant, tendant vers une « limite » au-delà de laquelle se place un spectre « continu ». On désigne ces raies sous le nom de *séries de Balmer.*

Dans l'infrarouge et dans l'ultraviolet lointain, on découvre d'autres raies en séries ayant exactement la même structure que la série de Balmer.

Ce que nous venons de dire des raies de l'hydrogène se retrouve, sensiblement identique, pour les raies émises par les autres atomes. Mais le schéma est d'autant plus complexe, les séries sont d'autant plus nombreuses que la masse atomique de l'élément qui produit la lumière est plus élevée.



La physique moderne a bien modifié l'ancienne conception de l'émission de la lumière par un atome. La description qu'en donne la théorie quantique (voir tome II) est fort différente de celle des « oscillateurs harmoniques » de la physique du XIXe siècle. De plus, elle est beaucoup plus complète. Elle permet d'interpréter l'infinie variété des spectres fournis par les éléments connus. Néanmoins, il y a un point commun entre les deux théories : c'est que le responsable de l'émission ou de l'absorption de lumière n'est autre que le mouvement d'un électron dans l'atome. Dans la théorie classique, un électron, en tournant régulièrement autour de l'édifice complexe que représente l'atome, joue le rôle d'oscillateur. Dans la théorie quantique, on ne considère plus les oscillateurs. On admet que l'électron émet de la lumière quand il « saute » d'un étage à l'autre de l'édifice atomique. A chacun des sauts possibles correspond l'émission ou l'absorption d'une raie. Ici ont été représentés les sauts qui correspondent aux premières raies des séries de Lyman (Lα, Lβ, etc), de Balmer (Hα Hβ, etc.) et de Paschen (Pα, Pβ, etc.) pour l'atome d'hydrogène.

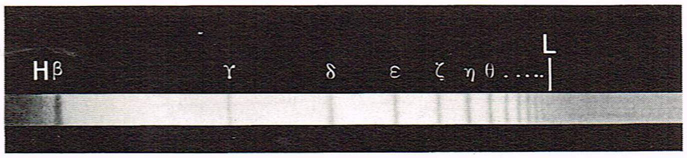
spectre et température

Observons, par une nuit claire et sans lune, la voûte céleste. Nous y voyons scintiller des étoiles qui diffèrent entre elles non seulement par l'éclat, mais encore par la couleur. On a cru, au XIXe siècle, que ces inégalités résultaient d'une diversité de la nature chimique des astres. Cette interprétation s'est révélée inexacte. On a montré, depuis, que la couleur des étoiles dépend simplement de leur température. Une expérience banale nous permettra de nous en rendre compte. Chauffons un fragment de métal ou de charbon; nous le verrons passer au rouge sombre, puis au rouge vif, et enfin, vers 1 500 °C, au blanc. Les diverses colorations des étoiles relèvent du même phénomène.

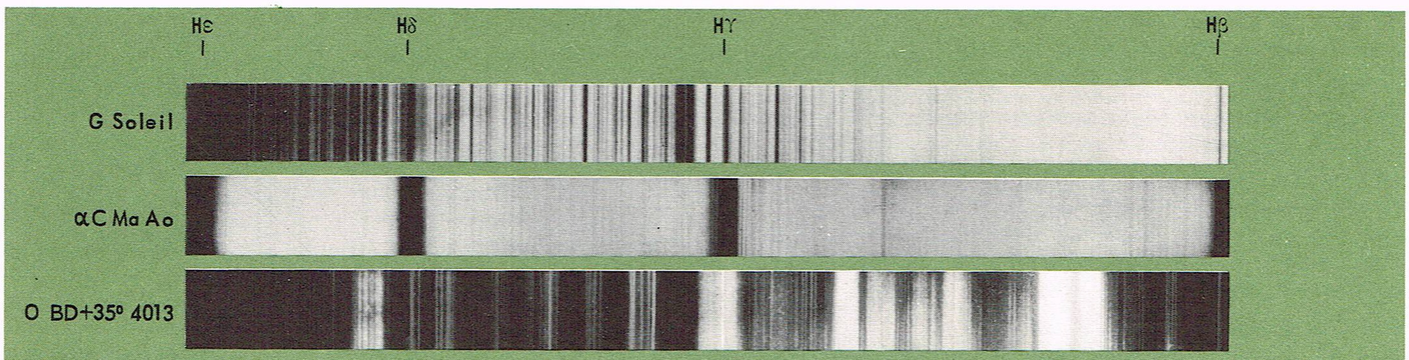
Nous avons tout d'abord des étoiles dites « froides », dont la température est inférieure à 1 000 °C. Elles sont de couleur rouge sombre. Pratiquement, elles sont invisibles, mais on sait qu'il en existe.

Viennent ensuite des étoiles « jaunes », de l'espèce de notre Soleil. Leur température est d'environ 6 000 °C à la surface. Beaucoup d'étoiles sont notablement plus chaudes qu'elles.

Vers 10 000 °C se placent les étoiles qui nous paraissent les plus brillantes. Tel est, par exemple, Sirius, qui se fait remarquer par sa belle lumière « blanche ». Enfin, les étoiles « bleues » sont encore plus chaudes. La limite de la température



La série de Balmer de l'hydrogène dans le spectre de l'étoile Véga (α Lyre).



Pour qu'un atome émette une ligne brillante ou absorbe des parties du spectre, ce qui y provoque des raies noires, il faut que ses électrons subissent une « excitation », c'est-à-dire une certaine modification, une sorte de déplacement au sein de l'édifice atomique. Cette excitation n'apparaît qu'à partir d'une certaine température, qui est au minimum de quelques milliers de degrés. Mais quand la température s'élève beaucoup plus — par exemple au-delà de 20 000 °C pour l'hydrogène —, on voit se produire un phénomène nouveau : l'électron n'est plus simplement déplacé au sein de l'édifice atomique; il est arraché à l'atome. On dit alors qu'il y a « ionisation ». L'émission de la série de raies spectrales n'est plus possible.
Nous avons ici un critère très sensible de la température.
Étoile G : 6 000° l'excitation de l'hydrogène est insuffisante.
Étoile A : 10 000° l'excitation est maximale.
Étoile O : 30 000° l'hydrogène est ionisé. Les raies de Balmer sont très faibles.
Seule, l'étoile A présente les raies de l'hydrogène d'une façon intense.

semble se situer vers 50 000 °C à la surface de l'astre.

La couleur donne une indication grossière de la température d'un astre. Cette indication a besoin d'être précisée par l'emploi de filtres colorés, qui permettent de comparer les températures des étoiles, deux à deux.

Une étoile qui émet principalement dans le bleu et dans l'ultraviolet sera défavorisée, en quelque sorte, par l'interposition d'un écran jaune : elle paraîtra très faible.

C'est là le principe d'une classification rapide des étoiles, qui se développe beaucoup actuellement.

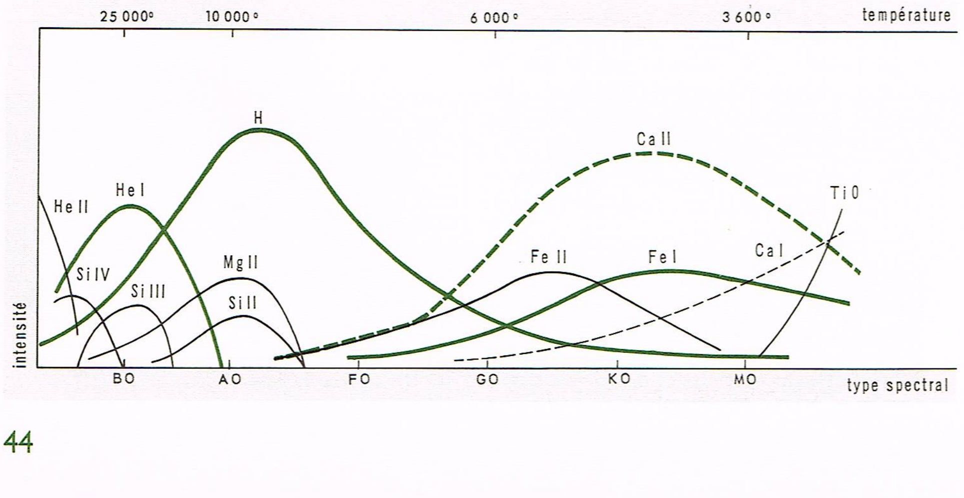
Mais, ainsi qu'il a été dit plus haut, la couleur d'une étoile ne nous donne qu'une indication grossière sur sa température. Une étude approfondie du spectre permet seule de préciser cette dernière : *c'est l'apparition, ou la disparition, de certaines séries de raies spectrales qui est le meilleur critère de température.*

spectre et composition chimique

Nous abordons un des problèmes les plus délicats de l'astrophysique : connaissant le spectre d'une étoile, pouvons-nous déterminer sa composition chimique? A vrai dire, ce problème est loin d'être résolu, bien que Fraunhofer, au XIXe siècle, ait fourni les grandes lignes d'une méthode. Il a identifié le spectre de l'hydrogène et d'autres éléments courants. Mais nous venons de voir que les raies spectrales caractéristiques d'un élément n'apparaissent que dans certaines limites de température, et qu'au-delà elles n'apparaissent plus, parce que l'excitation n'est plus réalisée. C'est ainsi qu'un ballon plein d'hydrogène n'émet pas de raies à la température ambiante.

En conséquence, l'absence des raies d'un élément ne prouve nullement que cet élément n'existe pas dans une étoile. De même, la prépondérance des raies d'un élément ne prouve pas que cet élément est abondant, mais seulement que les conditions thermiques de l'étoile favorisent la production de raies.

Le problème de la détermination des « abondances » des éléments dans une étoile donnée s'avère donc extrêmement délicat et complexe.



Intensité des lignes d'absorption en fonction du type spectral.

(D'après L. H. Aller.)

La tendance actuelle est de penser que toutes les étoiles ont pratiquement la même composition et que leurs dissemblances d'aspect ou de spectre proviennent des conditions d'ordre physique (notamment température et pression).

Un exemple est particulièrement typique : c'est celui des étoiles dites « à raies métalliques ». Leur spectre présente avec intensité les raies de certains métaux. Or, nul ne peut affirmer, à l'heure actuelle, que ces étoiles sont plus riches en métaux que les étoiles normales. Il est très possible que leurs atmosphères présentent des conditions physiques particulières qui suffisent à expliquer leur spectre.

Quelle est cette composition chimique des étoiles que l'on croit relativement semblable en tous les points de l'univers? Il semble qu'il s'agisse d'une sorte de matière universelle qui tantôt se condenserait pour former les étoiles, tantôt resterait diffuse sous forme de nébuleuses, tantôt fournirait la matière dispersée désignée sous le nom de « matière interstellaire ». Dans celle-ci, l'hydrogène est largement prépondérant. Pour 10 000 atomes d'hydrogène, on aurait au maximum 2 000 atomes d'hélium, 10 atomes d'oxygène, 2 de carbone et d'azote, etc. Citons encore quelques éléments qui existent, dans cette matière, à l'état de traces, c'est-à-dire en quantité très petites, mais connues; ce sont le sodium, le magnésium, le calcium, l'aluminium, le silicium et le fer.

spectre et mouvement stellaire

Parler du mouvement des étoiles, cela semble tout d'abord un paradoxe. Traditionnellement, les étoiles ne sont-elles pas les seuls objets fixes de l'univers?

Le Soleil, étant une étoile, fut également supposé fixe et, lorsqu'en mécanique on eut besoin d'un repère absolu, on prit le centre du Soleil comme origine. Mais, quand on regarde de plus près, cette fixité est extrêmement précaire. On a pu réaliser un film, accéléré un milliard de fois, de la voûte céleste : on y voit un ballet désordonné de toutes les étoiles prétendues fixes ... à l'exception de celles qui sont très éloignées et qui ne paraissent immobiles qu'en raison de la distance qui les sépare de nous.

Tout est mouvement dans l'univers. La Terre, comme chacun sait, tourne autour de son axe, et, de plus, elle tourne autour du Soleil à la vitesse énorme de 30 km/s, vitesse supérieure à celle des plus puissantes fusées de notre astronautique.

Mais le Soleil, lui aussi, se déplace. Avec toutes les étoiles qui nous environnent et qui constituent notre Galaxie, le Soleil tourne autour d'un point fixe de l'espace. La vitesse de ce mouvement est encore bien plus vertigineuse que celle de la Terre relativement au Soleil. On l'estime à 250 km/s au niveau du Soleil. Enfin, la Galaxie entière tourne sur elle-même et décrit un ballet dans l'espace.

Comment s'y reconnaître dans un tel enchevêtrement de vitesses dont les effets s'additionnent pour nous Terriens?

Ici encore, l'étude des spectres apporte une solution presque miraculeuse, et ce miracle c'est l'effet Doppler-Fizeau, plus amplement décrit au tome II.

Lorsqu'une source de lumière (étoile) se rapproche de l'observateur, la « fréquence » de sa lumière paraît plus grande, et il en résulte une diminution de la longueur d'onde observée. En conséquence, dans le spectre, les raies semblent se déplacer vers le bleu (puisque les longueurs d'onde vont en diminuant du rouge au violet). Lorsque, au contraire, l'astre s'éloigne de nous, les raies de son spectre paraissent déplacées vers le rouge.

Nous avons ici un moyen extrêmement simple et précis de mesurer la vitesse d'un astre par rapport à la Terre : vitesse que l'on dit « relative », car il est impossible de savoir si c'est l'astre qui se meut dans la direction de la Terre, ou la Terre dans la direction de l'astre. On constate finalement que toutes les étoiles sont en mouvement. Leurs vitesses s'échelonnent entre quelques kilomètres à la seconde et quelques dizaines de kilomètres. Elles peuvent atteindre exceptionnellement 100 ou même 200 km/s.

Ce même phénomène nous permet aussi d'obtenir des renseignements sur les mouvements internes des étoiles. On appelle « étoiles pulsantes » des étoiles qui se gonflent et se contractent alternativement, selon un rythme parfaitement régulier. L'effet Doppler-Fizeau nous permet de mesurer la vitesse de cette expansion et de cette contraction pendant les deux phases.

Une des applications les plus précieuses, pour le théoricien, de l'effet Doppler concerne les mouvements intenses des gaz qui composent ou entourent une étoile. Toute particule gazeuse est animée d'un mouvement rapide et désordonné, connu sous le nom d' « agitation thermique ». Ce mouvement est d'autant plus violent que la température est plus haute. Il se traduit par un élargissement des raies d'absorption du spectre.

Concluons : grâce au phénomène appelé « effet Doppler », l'étude méticuleuse du spectre permet de préciser les divers mouvements d'une étoile : déplacement d'ensemble, rotation, expansion et contraction, et mouvement des parcelles de matière au sein de l'atmosphère stellaire.

Rassemblons les différentes notions qui viennent d'être exposées : **le spectre est un puissant message de « surface » que nous adresse une étoile. Ainsi que nous l'avons montré, il peut nous servir de thermomètre, il permet une analyse chimique et il donne une idée des mouvements de l'astre. Ajoutons que l'étude de la largeur des raies spectrales permet d'estimer la pression gazeuse à la surface de l'étoile, et qu'elle fournit des indications sur le champ magnétique ainsi que sur l'agitation ou « turbulence » des atomes gazeux. Enfin, cette étude est le seul moyen de reconnaître si une étoile est simple, ou si elle est formée de deux ou trois étoiles très proches, tournant les unes autour des autres si serrées qu'on ne les distingue pas au télescope. De tels systèmes doubles ou triples sont très fréquents.**

Mais nous ne devons pas oublier que ce message, si important et si riche soit-il, n'est qu'un message de « surface ». Nous n'avons aucun moyen de pénétrer au sein de l'étoile, de savoir ce qui se passe, par exemple, en son centre.

Cependant, on peut arriver à avoir quelque idée de l'intérieur des astres en partant des conditions de surface, sur lesquelles le spectre nous informe, et en appliquant les lois de la physique terrestre. Tel sera précisément l'objet du prochain chapitre.

quelques objets étranges de l'astronomie

Nous avons déjà parlé, à propos de la détermination des distances astronomiques, des étoiles à éclat régulièrement variable : RR de la Lyre et céphéides.

On découvre dans le ciel des objets dont les propriétés lumineuses sont encore bien plus étranges.

C'est ainsi que, dans la Nébuleuse d'Orion, des étoiles dites « explosives » présentent des éruptions très brèves et qui se reproduisent à intervalles réguliers, au cours desquelles leur éclat augmente très sensiblement.

Un autre phénomène curieux est l'apparition d'une nova. Ce sont des étoiles qui deviennent brusquement brillantes, leur augmentation d'éclat atteignant en quelques heures 10 à 12 magnitudes.

On a même découvert des étoiles explosives dont l'éclat maximal est beaucoup plus élevé que celui des novae. On en a observé trois dans notre Galaxie; la première a été vue l'année 1054, la deuxième en 1572 par Tycho Brahé, la troisième en 1604 par Kepler.

On les désigne par le nom de « supernovae ». On admet qu'il n'en apparaît en moyenne qu'une seulement dans chaque galaxie, tous les 500 ans. La première supernova dont les traces restent marquées dans l'histoire date donc de 1054. Après son explosion, sa matière s'est répandue dans l'espace, mais en restant tout de même localisée. Elle forme actuellement une nébulosité, la nébuleuse du Crabe, qui montre d'étranges propriétés : c'est une source d'émissions radio-électriques intenses; sa lumière visible, entièrement polarisée, ne s'explique pas par les lois simples qui régissent les autres astres (on pense qu'elle est émise par des électrons tournant à grande vitesse dans des champs magnétiques). Cette nébuleuse est une intense radiosource, que nous retrouverons en radio-astronomie.

Il est encore d'autres étoiles curieuses, bien que moins rares et moins spectaculaires que les supernovae. A l'opposé des étoiles éruptives, on connaît une quarantaine d'étoiles, connues sous le nom de « R de la Couronne boréale », qui, à intervalles réguliers, deviennent brusquement plus faibles. On décèle dans leur spectre les raies du carbone et un nombre élevé de raies métalliques. Cette abondance du carbone amène à penser que ces étoiles rejetteraient brusquement de temps à autre une couche de gaz, dans lequel le carbone en abondance se condenserait en flocons de suie. Il se formerait un nuage opaque qui cacherait l'étoile, provoquant un obscurcissement temporaire; en s'éloignant de celle-ci, il redeviendrait peu à peu transparent, et l'étoile reparaîtrait avec son éclat primitif.

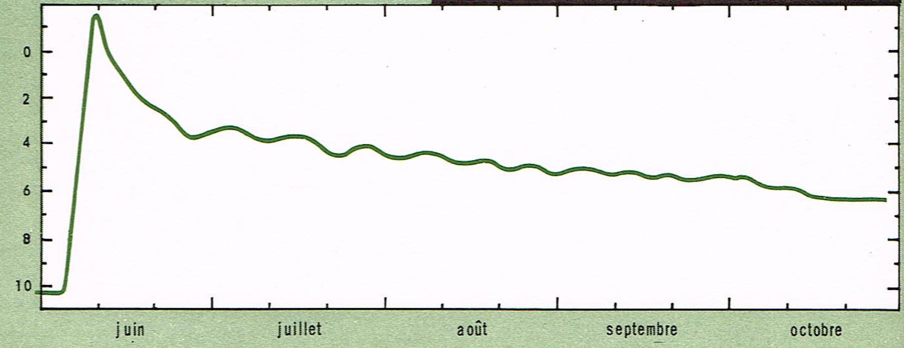


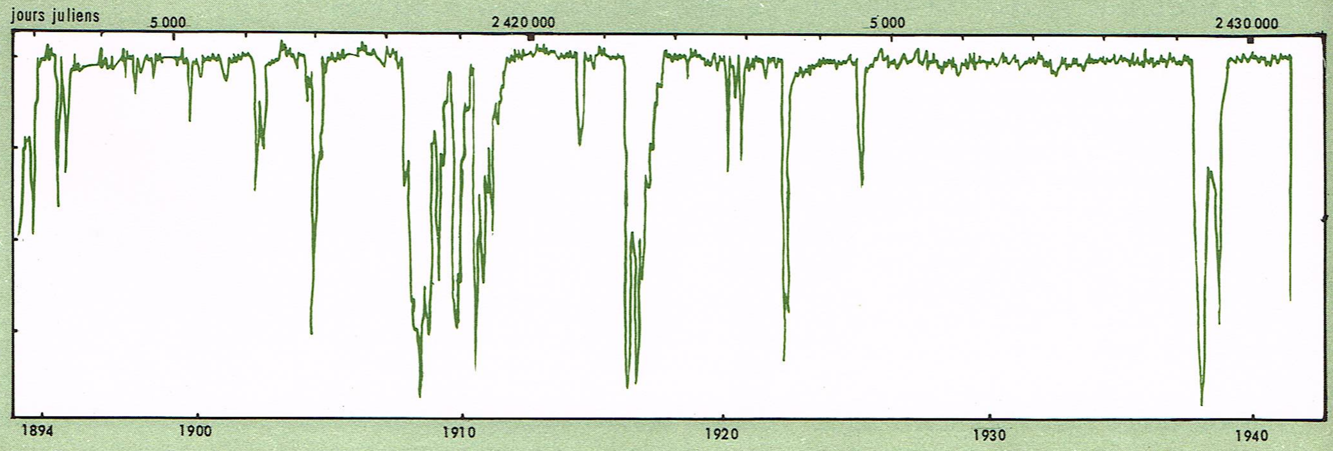
La nova de l'Aigle de 1918.

En haut, développement progressif d'une nébuleuse autour de l'étoile. 1. 20 juillet 1922; 2. 3 septembre 1926; 3. 14 août 1931.

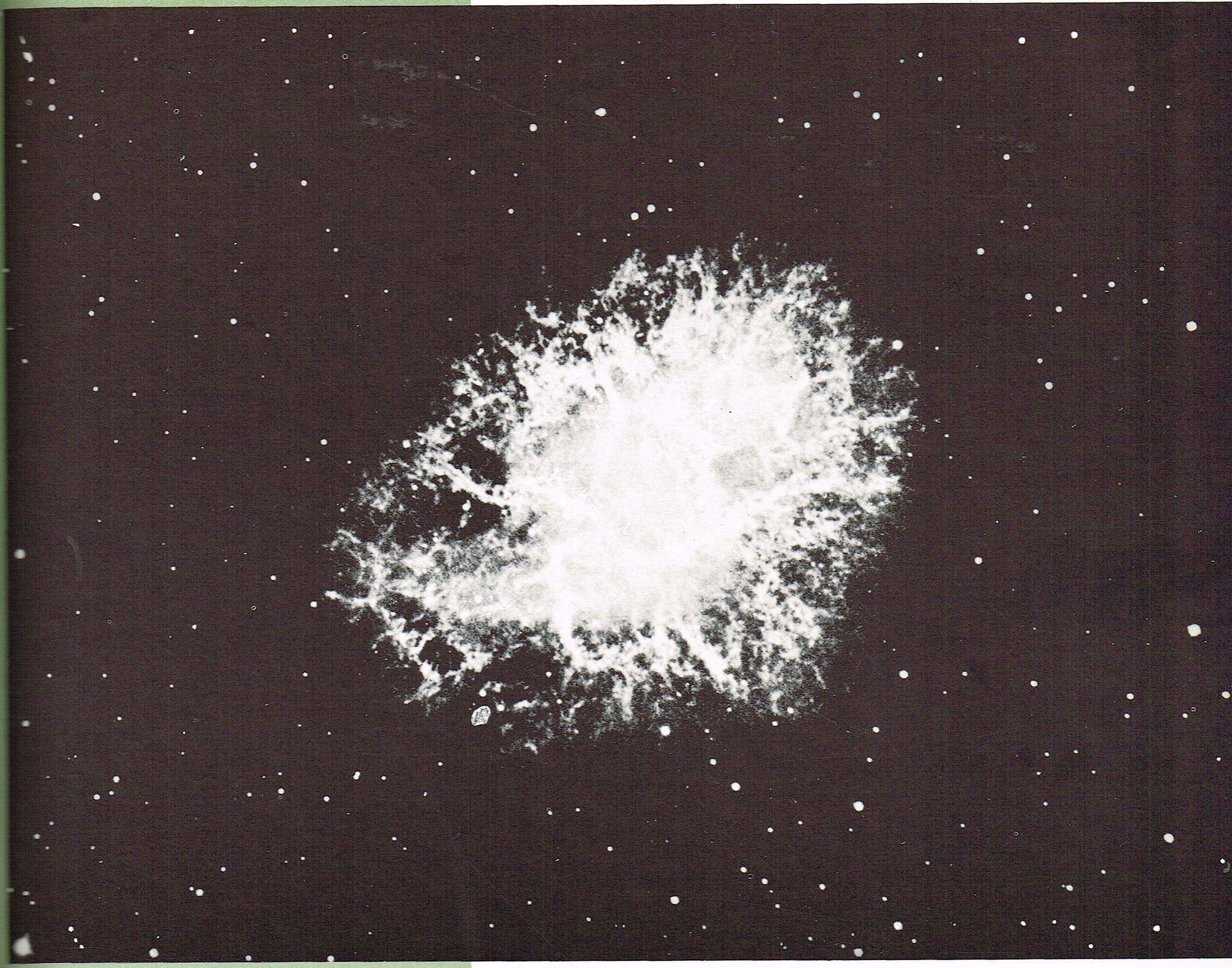
Ci-contre, courbe de lumière. L'étoile est passée en trois jours, du 7 au 10 juin, de la onzième grandeur à la magnitude — 1,5.

Ci-dessous, courbe de lumière de la variable R Couronne Boréale de 1894 à 1942.





Phot. Observatoire du Mont-Wilson.



La nébuleuse du Crabe, développement d'une supernova, observée en 1054.

Phot. W. Baade (observ. du Mont-Wilson).

peut-on classer les étoiles ?

On classa d'abord les étoiles par constellations, à qui les Anciens donnèrent les noms les plus poétiques. Bien entendu, de tels groupements entre astres situés dans la même direction de la voûte céleste, mais non nécessairement à la même distance de nous (donc, finalement, dispersés au hasard) et n'ayant pas les mêmes propriétés, sont aussi artificiels et vides de sens physique que ceux qu'on obtiendrait en classant les accidents géographiques selon leur direction sur une table d'orientation.

Intervint ensuite une classification par « grandeur » (éclat), presque aussi artificielle, car la magnitude d'une étoile dépend tout autant de son éloignement que de sa luminosité.

On apprit ensuite à classer les étoiles par couleur : la couleur est, en effet, indépendante de l'éloignement et de la position de l'observateur, à la différence des précédentes; *on dit qu'elle est « intrinsèque », ce qui est une vertu inappréciable pour une donnée d'observation.* Avec les progrès de la spectroscopie s'imposèrent ensuite les notions de type spectral et de classe de luminosité.

Avant de voir en détail ce que tout cela signifie, et sur quels immenses travaux sont bâtis les systèmes actuels, disons tout de suite que seule l'idée d'évolution et de filiation génétique donne son plein sens à toute classification stellaire.

Les classifications actuelles des étoiles utilisent des critères différents, suivant la finesse et la puissance de l'instrument employé. C'est la raison pour laquelle chaque groupe d'observateurs a ses propres critères d'observation. On parle ainsi des classifications de Harvard, de Yerkes et de l'Institut d'astrophysique de Paris.

les types spectraux classification de Harvard

Il y a actuellement 400 000 étoiles classées selon leur type spectral. On utilise la nomenclature suivante : O B A F G K M, allant de ce qu'on appelle les « premiers types », qui correspondent aux étoiles les plus chaudes, aux « types avancés », qui correspondent aux plus froides. Il paraîtrait plus logique de classer les étoiles en respectant l'ordre alphabétique. C'est précisément ce qui fut fait, mais à une époque où la corrélation entre l'aspect d'une étoile et sa température était encore mal connue; quand elle fut totalement élucidée, on s'aperçut qu'il était nécessaire de permuter l'ordre de certains types spectraux et d'en abandonner d'autres (C, D, E, etc.). Chaque classe est divisée en dixièmes de classe, par exemple B0, B1, B2... Ainsi, Sirius est une étoile du type A2; notre Soleil est du type G1; etc. Cette succession régulière de types spectraux est en fait une « séquence de température », puisque finalement les raies n'apparaissent que dans un domaine relativement restreint de température.

La classification dont nous venons d'ébaucher les grandes lignes est essentiellement l'œuvre de l'observatoire de Harvard. On dit que c'est une *classification à un seul paramètre*. En effet, les types spectraux y sont uniquement fonction de la température des étoiles.

classes de luminosité

Avec les progrès de la spectroscopie, et aussi des études théoriques concernant les atmosphères stellaires, on s'aperçut que cet unique paramètre était insuffisant pour caractériser entièrement les étoiles. Un autre facteur essentiel intervient, c'est la pression qui règne dans l'atmosphère : à une forte pression correspondent des raies larges, diffuses; à une pression faible, des raies fines et bien nettes. On a été amené ainsi à introduire un nouveau paramètre : la *classe de luminosité*.

La classification à deux paramètres, dont nous allons parler maintenant, est essentiellement l'œuvre de Morgan et de Keenan, de l'observatoire Yerkes. On en comprendra facilement l'idée directrice. Soit, à masse et à température égales, deux étoiles, dont l'une se trouve à une pression beaucoup plus faible que l'autre : donc, nécessairement, son gaz occupe un volume beaucoup plus grand. Une telle étoile est appelée « géante », et, lorsque ce phénomène est encore plus marqué, c'est une « supergéante ».

En revanche, et par opposition, les étoiles normales sont qualifiées de « naines ». On conçoit qu'à température égale les géantes rayonnent beaucoup plus d'énergie.

Ainsi, une étoile géante dont la température de surface est de l'ordre de 4 000 °C aura environ 10 magnitudes de moins (sera 100 fois plus brillante) qu'une étoile naine de même température. Une étoile supergéante comme Bételgeuse a sensiblement 16 magnitudes de moins que l'étoile naine correspondante, de type M3. On calcule facilement que son rayon doit être plusieurs centaines de fois plus grand.

Parmi ces astres si différents, il semble qu'il y ait certains écarts de composition chimique, provenant de leur différence d'âges. **Ainsi voyons-nous apparaître l'idée d'une évolution, fil conducteur de toute classification vraiment rationnelle.** Ce point est tellement important que nous y reviendrons longuement plus loin.

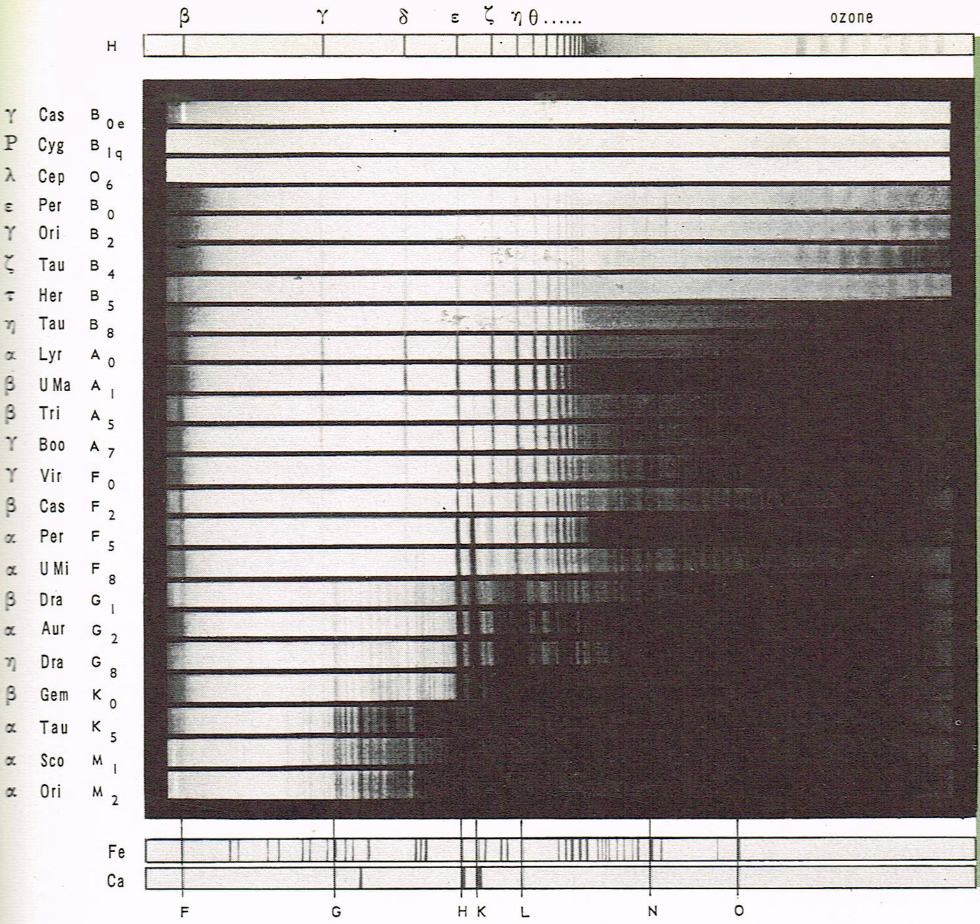
les classifications intrinsèques

La classification de Harvard en types spectraux et son perfectionnement par l'équipe de Yerkes utilisent des caractères empiriques, souvent difficiles à relier aux véritables caractères physiques des étoiles. De plus, ces critères sont discontinus, ce qui leur confère quelque chose d'irrationnel. Telle raie fera classer l'étoile dans l'une des sous-classes de A0 à A3, telle autre de F1 à F4, etc. Ce défaut est inévitable quand on utilise comme critères de classification l'apparition et la disparition des raies dans des domaines de température étroits.

C'est pourquoi on a cherché des critères qui soient plus directement liés à l'état physique des atmosphères stellaires (température et pression) et qui restent utilisables pour tout l'ensemble de la séquence spectrale.

Cette grande synthèse, qui atteint actuellement le plus haut degré de perfection, est l'œuvre de l'Institut d'astrophysique de Paris (Barbier et Chalonge).

Indiquons-en simplement le principe : au lieu d'étudier les raies du spectre stellaire, on étudie systématiquement, par des moyens photométriques extrêmement précis, la répartition de l'énergie dans le fond continu du spectre de l'étoile.



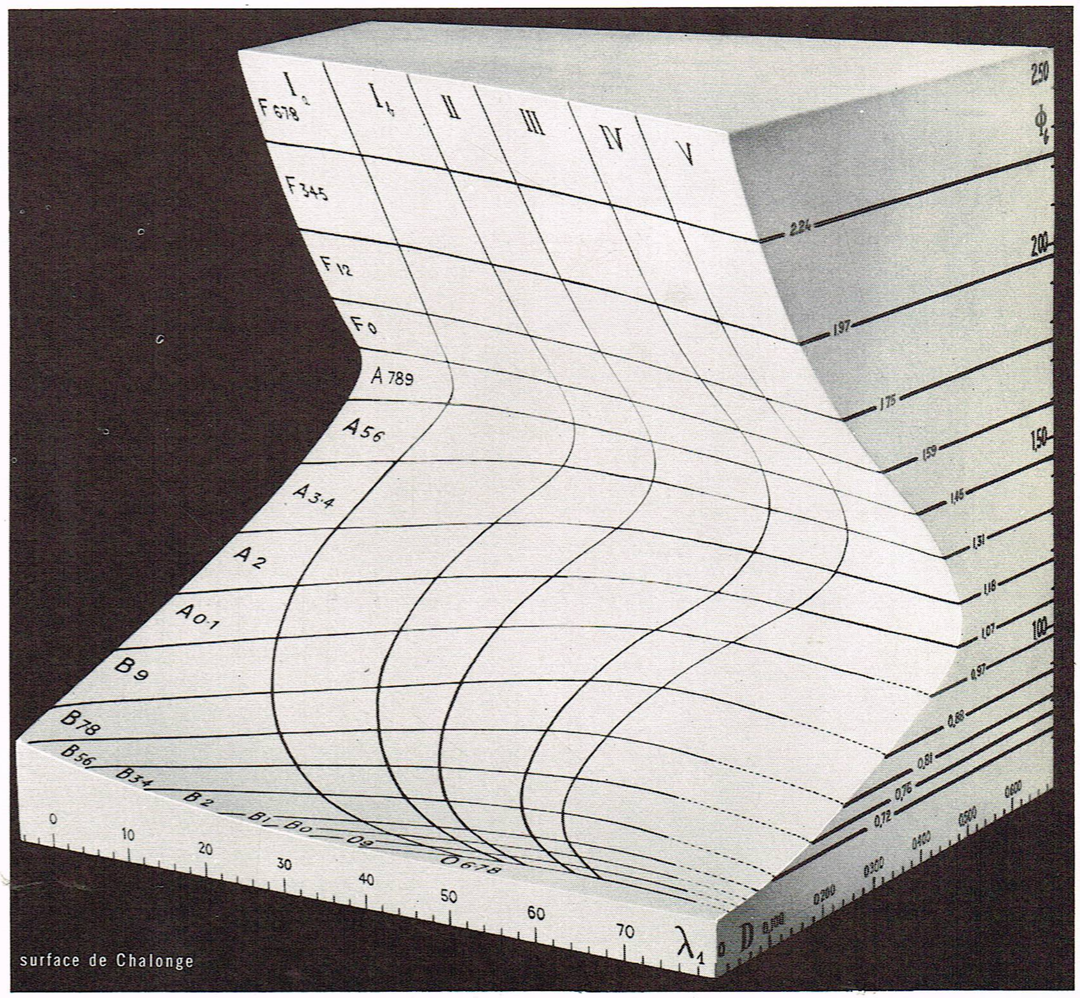
Série de spectres stellaires. Les caractères dominants de chaque type sont indiqués ci-dessous :

Type	*Température*	*Caractères dominants*
O	50 000°	Raies de l'hydrogène faibles Raies de l'hélium ionisé
B	15 000°	Raies de l'hélium ionisé fortes Raies de l'hydrogène plus marquées
A	10 000°	L'intensité des raies de Balmer passe par son maximum pour ce type spectral. Apparition des raies du calcium ionisé (Ca II) [ex., Sirius, Véga]
F	7 000° à 8 000°	Raies du calcium ionisé très intenses Raies de l'hydrogène encore visibles (ex., Procyon)
G	6 000°	Spectre du type solaire Raies du calcium ionisé au maximum d'intensité. Spectres très riches en raies métalliques
K	5 000°	Raies des atomes neutres (ex., Aldébaran)
M		Raie du calcium neutre Apparition de bandes moléculaires d'oxyde de titane (ex., Bételgeuse)

En principe, il suffit de chercher quelle serait la température d'un corps chauffé à l'incandescence (corps noir) qui donnerait cette même répartition, pour avoir la température, donc le type spectral, de l'étoile. Cette comparaison se fait au moyen de ce qu'on appelle les « gradients spectro-photométriques ». Nous ne pouvons pas insister sur ces techniques, complexes certes, mais qui reposent sur un support théorique absolument simple et logique.

La détermination de la « classe de luminosité » se fait suivant les mêmes principes. Plus les raies sont fines, plus grand est le nombre des raies de Balmer qu'on peut distinguer vers la limite de convergence de la série. Dans certaines étoiles supergéantes, on distingue jusqu'à 33 raies de Balmer, alors que, pour une étoile naine, on n'en distingue qu'une dizaine.

Un dernier raffinement de la technique de l'Institut d'astrophysique de Paris vise à éliminer les modifications du spectre produites par les masses de « matière interstellaire » interposées entre l'étoile et nous, et qui risquent de fausser le classement de l'astre. Ce troisième paramètre est une grandeur absolument indépendante de ces actions interstellaires : la discontinuité de Balmer. Nous avons expliqué ce que sont les raies de Balmer. A la limite de la série qu'elles forment, on



surface de Chalonge

constate un brusque affaiblissement de la luminosité du spectre, d'autant plus marqué que les raies de Balmer sont plus intenses. On l'explique par une absorption « continue » au-delà de la limite de Balmer.

Si l'on classe les étoiles au moyen de ces trois paramètres, pris comme coordonnées, on constate que la plupart d'entre elles se placent sur une surface, la surface de Chalonge, dont nous donnons la reproduction à la page précédente.

Comme on vient de le voir, les astronomes sont passés, pour la classification des étoiles, de critères empiriques à des critères physiques et rationnels.

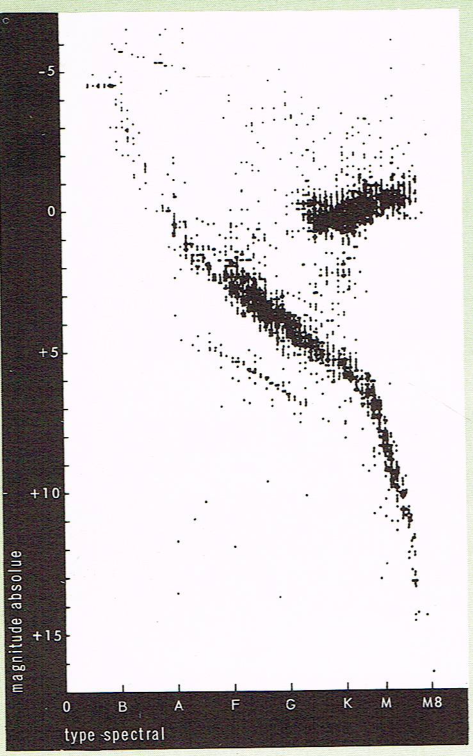


Diagramme de Hertzsprung-Russell.

(D'après Parenago.)

le diagramme de Hertzsprung-Russell, carrefour et clef de voûte de l'astronomie

La luminosité des étoiles présente un « éventail » de variations extrêmement grand. Certaines sont un million de fois plus lumineuses que le Soleil, d'autres un million de fois moins.

Les températures « de surface » des étoiles les plus chaudes sont de l'ordre de 100 000 °C; celles des plus froides, de l'ordre de 1 000 °C. (Il en existe de plus froides encore, mais nous ne pouvons les voir.)

Dès les premières années de ce siècle, Hertzsprung et Russell remarquèrent que, dans l'ensemble, les étoiles très lumineuses émettent une lumière d'un blanc bleuté, tandis que les étoiles peu lumineuses donnent une lumière rouge-orangé. Nous dirons que les étoiles très lumineuses appartiennent plutôt aux premiers types spectraux (O, B) et les étoiles peu lumineuses aux types dits « avancés » (G, K, etc.).

En 1911, Hertzsprung porta en ordonnées la luminosité (exprimée en magnitude absolue) et en abscisses le type spectral, c'est-à-dire pratiquement la température de surface de chaque étoile. Il constata alors que les étoiles, loin de s'éparpiller au hasard dans ce diagramme, s'assemblent autour d'une ligne courbe qui traverse le plan en diagonale. Cette région a été appelée la « séquence principale ». Après séparation nette de ce groupe, on trouve un second groupe plus limité d'étoiles, en haut et à droite du diagramme; ce sont les géantes et les supergéantes.

La détermination du diagramme de Hertzsprung-Russell pose le problème des distances. En effet, pour l'établir, on porte en ordonnées la luminosité réelle de l'étoile et non sa magnitude apparente, qui dépend essentiellement de son éloignement par rapport à nous. Mais le rôle de tels diagrammes est si essentiel et central qu'ils dominent les études concernant les problèmes d'évolution stellaire auxquelles seront consacrés les chapitres suivants. Montrons simplement ici qu'il permet d'intéressantes estimations des distances stellaires : en effet, si une étoile a un type spectral parfaitement bien déterminé, nous pouvons déterminer sa magnitude absolue sur le diagramme de Hertzsprung-Russell. La différence entre sa magnitude apparente et sa magnitude absolue permet d'estimer sa distance. C'est ce qu'on appelle la « méthode des parallaxes spectroscopiques », dont nous avons dit un mot au premier chapitre. Elle s'applique à des distances supérieures à celles des parallaxes trigonométriques, mais ne permet pas d'aller aussi loin que les méthodes qui utilisent les étoiles variables comme indicateurs de distance.

Comme on vient de le voir, l'astronomie moderne ne s'est plus contentée de classer les étoiles selon leur éclat apparent, ou grandeur, et leur coloration. Grâce à la spectrographie, elle a pu accumuler un nombre considérable de données sur l'état physique et la composition chimique des corps célestes. Il en résulte des classifications qui font appel à plusieurs caractères; nous verrons même qu'elles permettent de déterminer l'âge des étoiles. **Ainsi, comme la classification des êtres vivants, celle des étoiles débouche sur l'idée d'évolution, qui en devient le fil directeur.**

étoiles doubles

Si l'on suppose les étoiles réparties au hasard dans l'espace, la probabilité pour que deux étoiles soient très voisines dans le champ du télescope est extrêmement faible. Or, l'observation nous montre, au contraire, un grand nombre d'étoiles groupées par deux; les systèmes multiples (trois et plus) sont même assez fréquents *(plus de la moitié des étoiles proches font partie de systèmes multiples)*. Une étude plus approfondie de ces groupements, et, en particulier, l'étude du mouvement des deux astres, montre avec certitude qu'il s'agit effectivement de « systèmes physiques » dans lesquels les deux étoiles sont proches dans l'espace, et non pas seulement vues dans la même direction. La simple observation visuelle avec un bon instrument permet la découverte d'un grand nombre d'étoiles doubles.

On en connaît des milliers. Lorsque le couple est trop serré pour que les deux étoiles puissent être distinguées, des méthodes physiques permettent de déceler tout de même ces groupements : par exemple, l'étude spectroscopique montre un spectre composite, mélange de ceux des deux étoiles. Parfois, aussi, l'une des étoiles est sensiblement moins brillante que l'autre, et la masque périodiquement : on dit qu'on a alors une « variable à éclipse ». Dès 1782, Goodricke expliquait de cette manière (à tort, car il s'agissait dans ce cas d'une étoile variable) les variations périodiques d'éclat d'Algol.

Un des caractères constants des systèmes doubles est leur mouvement « orbital ». Expliquons-nous : la Lune, comme les satellites artificiels, gravite indéfiniment autour de la Terre. On peut considérer que le mouvement de la Terre n'est pas modifié par le lancement d'un satellite, tant sa masse est petite.

Mais si nous lancions des satellites de plus en plus gros, atteignant la moitié de la Terre ou même plus, lequel graviterait autour de l'autre, lequel serait alors le « satellite »? Ni l'un ni l'autre : les deux solides tournent alors autour d'un point fixe, le centre de gravité des deux masses (v. tome II, chap. *La mécanique à l'heure spatiale*.)

Dans le cas des étoiles doubles visuelles, l'observation directe des orbites est possible. Au contraire, dans le cas des « doubles spectroscopiques », seul l'effet Doppler permet, par l'étude du déplacement régulier des raies dans le spectre, l'estimation de la vitesse radiale des étoiles au cours de leur giration.

L'étude de la trajectoire des étoiles doubles a une importance primordiale en astrophysique, car ces étoiles sont les seules pour lesquelles une détermination directe des masses soit possible. Les observations systématiques d'un grand nombre de couples d'étoiles laissent pressentir que la luminosité d'une étoile augmente avec sa masse. Eddington a émis l'hypothèse que toutes les étoiles de même éclat réel ont aussi la même masse. Mais l'éventail des masses connues est beaucoup plus faible que celui des luminosités : on ne connaît guère d'étoiles « pesant » plus de dix Soleils; à l'opposé, une masse dix fois plus faible semble être une limite inférieure. La luminosité varie d'une étoile à l'autre dans des proportions infiniment plus grandes. La relation entre la masse et la luminosité n'est donc pas une simple relation linéaire; la luminosité varie comme la quatrième puissance de la masse.

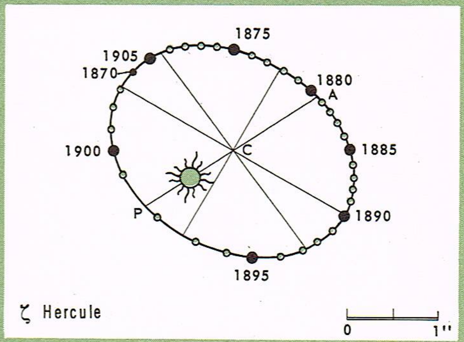
Les « variables à éclipse » fournissent aussi un autre renseignement précieux : les rayons des deux étoiles. En effet, la courbe de lumière permet de calculer les dimensions relatives des deux astres. Il est possible d'évaluer les rayons en valeurs absolues, c'est-à-dire en kilomètres, si l'on connaît le diamètre des orbites. Ce calcul est possible, car les mesures spectroscopiques du déplacement des raies par effet Doppler permettent de connaître la vitesse des astres. De la mesure de leur période de giration on déduit la longueur de l'orbite.

Masses, rayons, voici des données fondamentales dont les étoiles doubles nous fournissent une ample moisson. Et lorsque nous aurons dit qu'on trouve, parmi elles, des représentants de la plupart des types spectraux, et qu'on admet à l'heure actuelle que plus de la moitié des étoiles sont engagées dans des systèmes doubles et triples, on comprendra qu'un grand nombre d'astrophysiciens aient consacré et consacrent encore leur vie entière à l'étude de cette intéressante population.

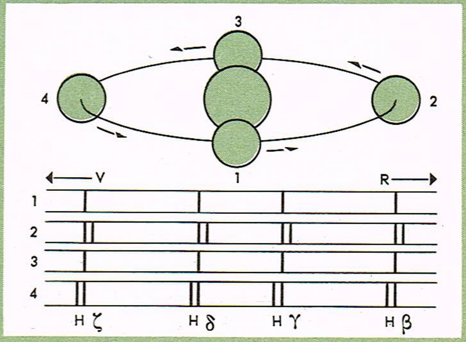


Les étoiles doubles serrées, c'est-à-dire dont les deux composantes sont très proches l'une de l'autre, présentent fréquemment une enveloppe gazeuse commune. Parfois même, on peut déceler de véritables jets de gaz d'une étoile à l'autre : un exemple remarquable est celui de l'étoile β de la Lyre.

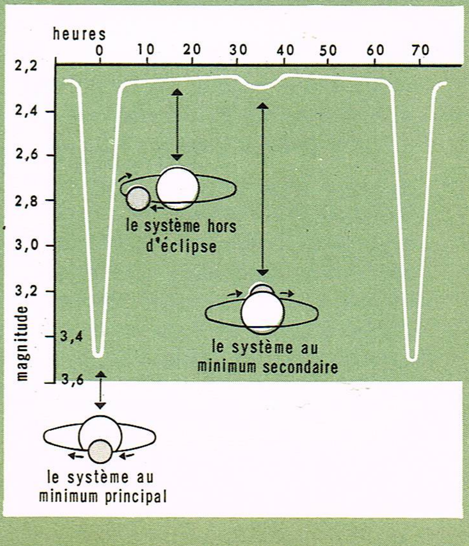
(D'après O. Struve et G. P. Kuiper.)



Les étoiles doubles physiques se reconnaissent au mouvement orbital de leurs composantes. Ici, ζ Hercule, période 35 ans.



Dédoublement des raies spectrales des étoiles doubles spectroscopiques.



Courbe de lumière d'une variable à éclipse.

peut-on atteindre le centre même des étoiles ?

On peut à juste raison penser qu'aucun sujet d'astrophysique ne nous est plus inaccessible que ce qui se passe au sein des étoiles : la « profondeur » à laquelle notre investigation pénètre dans l'atmosphère stellaire est d'une petitesse propre à décourager. Nous avons déjà vu qu'il n'est que rarement possible de déterminer la masse, le rayon et la luminosité des étoiles.

Comment peut-on espérer, à partir de ce maigre butin, esquisser une analyse de la structure interne de ces astres, deviner dans quel état y est la matière, quelles sont les pressions et les températures qui y règnent? Sont-elles gigantesques, et devons-nous imaginer des « états physiques » inconnus de nous (qui existent peut-être aussi dans la constitution interne de notre propre globe)?

Nous n'avons qu'une possibilité d'aborder ces mystères : admettre que les lois physiques de la matière sont les mêmes au sein des étoiles qu'à la surface et à l'intérieur de la Terre.

Cette attitude est-elle justifiée?
Remarquons tout d'abord qu'elle est la seule possible. On ne peut aborder un problème avec des instruments qu'on ne possède pas.

De plus, l'histoire a montré que la physique de l'infiniment grand est étroitement liée à notre physique terrestre et lui ressemble en tous points : les lois de la mécanique que nous appliquons à la chute des corps et à une infinité de problèmes de la vie courante sont celles mêmes qui régissent les mouvements des planètes du système solaire, et très probablement aussi ceux des astres lointains. Depuis longtemps on sait prévoir avec une certitude absolue l'apparition d'une éclipse et de beaucoup d'autres phénomènes astronomiques, la venue d'une comète par exemple; et ce phénomène, qui a terrorisé nos ancêtres, n'a plus guère de mystère pour nous.

Si nous nous intéressons maintenant à des questions de physique proprement dite, nous voyons toutes les grandes découvertes faites dans notre monde s'appliquer immédiatement à l'astrophysique. Parfois même c'est précisément des astres lointains que nous viennent les meilleurs arguments pour consolider les nouvelles théories. Et le royaume de l'infiniment petit n'est pas exclu de cette vaste symbiose. Pensons à la mécanique quantique, qui explique les émissions des spectres stellaires. Pensons aux rayons cosmiques, qui nous apportent ces particules étranges qu'avaient prévues les théoriciens, mais qu'ils ne savaient pas produire au laboratoire. Pensons enfin aux réactions « nucléaires », qui, avec l'intense énergie qu'elles dégagent, ont donné la réponse au problème, apparemment insoluble, de la génération d'énergie au sein des étoiles.

Il ne paraît donc pas insensé d'appliquer les lois de notre physique terrestre, locale, aux astres lointains. Nous ne voulons pourtant pas donner au lecteur une image trop optimiste; à l'heure actuelle, il serait présomptueux de parler d'un calcul rigoureux; en vérité, il ne s'agit que d'estimations raisonnables et sérieuses des conditions qui règnent dans ces régions. Il faut reconnaître que certains problèmes paraissent échapper à notre physique terrestre; ainsi, aucune explication n'a été donnée, à ce jour, au phénomène qu'on appelle l' « expansion de l'univers », sur lequel nous reviendrons dans un prochain chapitre. De même, il n'a pas encore été possible d'expliquer l'origine des champs magnétiques qui entourent planètes et étoiles, et qui règnent jusque dans l'espace interstellaire.

Nous allons maintenant montrer comment, avec les données d'observation que nous possédons sur les étoiles, nous pouvons essayer de « descendre » vers leur intérieur. Une remarque paraît ici s'imposer; elle concerne la constance de la structure des étoiles pendant de très longues périodes. L'étoile est normalement un système en « quasi-équilibre » physique.

Les observations astrophysiques systématiques ne remontent guère qu'à une cinquantaine d'années. C'est peu par rapport à la durée des « temps astronomiques », et il peut sembler imprudent d'accepter l'idée de la constance de la structure des étoiles pendant des millions ou des milliards d'années (1). De nombreuses observations appuient cependant cette thèse. On connaît, par exemple, l'existence d'algues fossiles dont l'âge est supérieur à un milliard d'années; on peut en conclure que la température de la Terre n'a pas varié de plus de vingt degrés depuis cette époque reculée, et cela implique la constance de la luminosité du Soleil pendant ce temps.

L'analyse de la structure des étoiles présuppose qu'elles sont dans un état d'équilibre parfait. Voyons d'abord quelles sont ces conditions d'équilibre, que nous appellerons nos « hypothèses de travail ». Ensuite nous indiquerons les conséquences qu'on en peut déduire.

quelques hypothèses de travail

Les structures stellaires doivent rester permanentes et stables pendant des millénaires; on parlera d'équilibre hydrostatique, d'équilibre mécanique, d'équilibre thermique, d'équilibre énergétique, etc.

Considérons d'abord la question sous un jour purement mécanique.

La matière stellaire est essentiellement fluide. Les températures qui y règnent, même en surface (de 3 000 °C à plusieurs millions de degrés), s'opposent à l'existence de tout corps solide. Il suffit d'écrire les conditions de l'équilibre hydrostatique pour être en état d'évaluer la dérivée de la pression : on conçoit que, de proche en proche, on puisse « descendre en profondeur » jusqu'au centre des étoiles. En faisant la somme de toutes les différences de pression, depuis le bord jusqu'à ce centre, on procède à ce qu'on appelle en mathématiques une « intégration ».

On peut ainsi estimer la pression au centre si l'on connaît la masse et le rayon de l'étoile. Elle est considérable, puisque les couches fluides placées au centre supportent le poids de toutes les couches supérieures.

Ainsi, au centre du Soleil règnent des pressions qui sont de l'ordre de plusieurs millions d'atmosphères.

(1) A l'exclusion des novae ou des supernovae, étoiles qui, à la suite d'une longue évolution, explosent subitement et subissent ainsi une véritable « mutation » les transformant, par exemple, en naines blanches.

De l'estimation de la pression nous pouvons passer d'emblée à celle de la température.

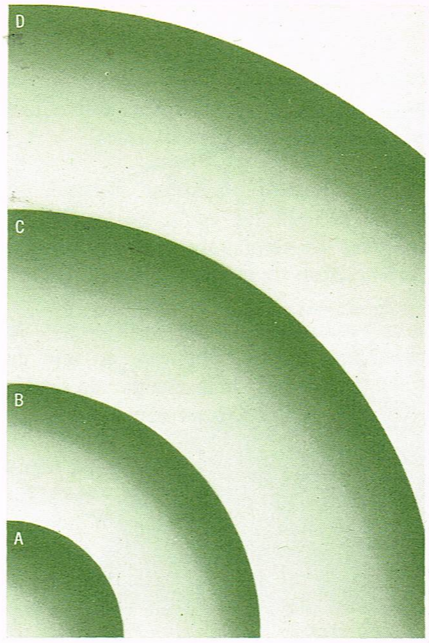
Dans les conditions usuelles, sur la Terre, des relations existent entre la pression, la température et la densité d'un gaz. Ces lois ($pv = rT$) s'appliquent bien aux gaz à faible densité, mais non aux pressions élevées ni aux basses températures, car il existe entre les molécules des interactions qui compliquent les phénomènes. On emploie alors des formules plus complexes, qui tiennent compte de ces interactions et du volume occupé par les molécules individuelles.

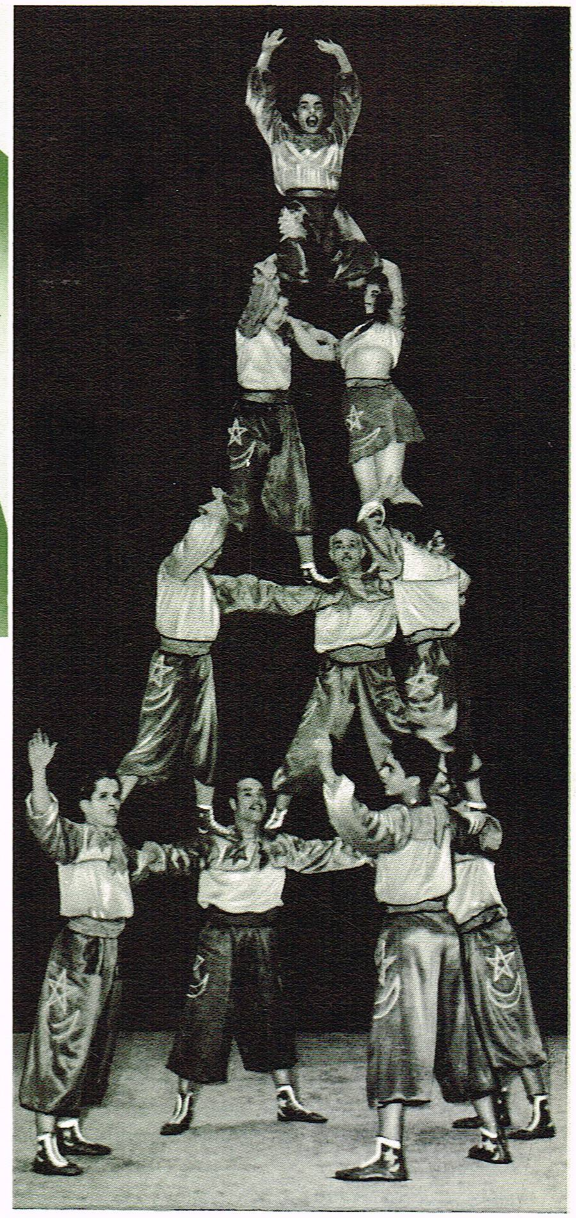
Dans les atmosphères stellaires, la densité est faible et la température élevée : ce sont, par excellence, les conditions où les gaz obéissent à des lois simples. Mais on peut se demander s'il en est de même à l'intérieur, où les pressions sont considérables. Ici intervient un facteur favorable : les températures que l'on peut prévoir sont si élevées que tous les atomes sont complètement ionisés, c'est-à-dire dépouillés de leurs électrons; leur volume effectif se trouve donc réduit à celui de leur noyau. Le volume des particules individuelles reste donc très petit par rapport au volume total du gaz, et les conditions semblent encore réunies pour que s'appliquent les lois des gaz parfaits ou celles des gaz dégénérés.

Les températures que l'on calcule ici sont d'un ordre de grandeur absolument inhabituel. Pour mieux nous en rendre compte, rappelons qu'au laboratoire on chauffe rarement des gaz à plus de 1 500 °K, et que les températures les plus élevées que nous sachions réaliser, celle d'un filament de lampe à incandescence, par exemple, sont de l'ordre de 3 000 °K.

Déjà dans les couches extérieures des étoiles nous avons affaire à un éventail de températures qui nous sont inconnues sur la Terre, et, pour étudier la physique de ces atmosphères, il nous faut « extrapoler » les lois usuelles jusqu'aux environs de 50 000 °K.

Au sein même des étoiles, nous rencontrons des températures d'un tout autre ordre de grandeur. Le calcul dont nous avons exposé le principe nous conduit à des températures de 10 millions de degrés. Il ne faut pas sous-estimer le « fossé » : les lois des gaz ne sont effectivement établies par l'expérience que jusqu'à 1 500 K°. Nous allons les appliquer à des millions de degrés (1).





La structure de l'étoile, en allant du centre au bord, pourrait être comparée à une troupe d'équilibristes, chacun étant soutenu par celui qui se trouve sous lui. Ce sont les équilibristes du bas (couche A) qui sont soumis à la plus forte pression, car ils supportent le poids des autres. Ce seront donc les athlètes les plus « costauds » qu'on placera en bas. Sur D, au contraire, ne s'exerce aucune pression, et D pourra être un jeune garçon.

Phot. Pic, Rapho.

Quelques certitudes cependant s'imposent, si nous voulons esquisser un tableau de cette fournaise. A ces températures, aucun corps composé ne peut subsister. Les éléments sont tous à l'état d'atomes isolés. Tout composé serait immédiatement dissocié. De plus, il est facile de concevoir qu'il n'existe ni corps solides ni liquides, mais uniquement des gaz. Enfin, nous l'avons dit, il est certain aussi que ces atomes sont entièrement ionisés à ces températures, et nous nous trouvons là devant un état physique encore mal connu : le *plasma* (v. plus loin, chap. *Les hautes températures*). Indiquons seulement ici que les estimations de température dont nous venons de donner une idée sont confirmées par plusieurs autres considérations qui se recoupent (taux de génération d'énergie nucléaire, etc.), et, même si les valeurs mises en jeu paraissent inhabituelles, les résultats ne sont nullement mis en question à l'heure actuelle.

les transferts d'énergie

Au point de vue énergétique, une étoile n'est pas dans un état d'équilibre au sens usuel du terme. Cette notion d'équilibre, si fructueuse dans la physique terrestre, implique que les diverses parties d'un système n'échangent pas d'énergie les unes avec les autres, ou que, si ces échanges

(1) La question se pose de la définition même de telles températures, qu'aucun instrument ne permet de mesurer. La théorie « cinétique » des gaz fournit la réponse : les particules qui constituent un gaz sont animées de mouvements désordonnés appelés « agitation thermique ». Leur vitesse d'agitation augmente avec la température. On convient de définir précisément cette dernière par la vitesse d'agitation des particules. C'est un aspect de l'équivalence entre l'énergie et la chaleur.

existent, ils s'annulent mutuellement. On dit que le « bilan net » des échanges d'énergie est nul.

La situation est tout autre au sein d'une étoile. Il faut imaginer au centre un noyau extrêmement chaud, qui émet une quantité énorme d'énergie sous forme de rayonnement. Cette énergie se propage progressivement vers l'extérieur de l'étoile, et l'on peut parler d'un véritable « flux » d'énergie qui s'écoule continuellement du centre vers la surface.

L'atmosphère stellaire, c'est-à-dire les minces couches extérieures qui sont responsables de la lumière que nous recevons des étoiles, a peu d'influence sur ce processus. Sa température s'ajuste de telle sorte que l'énergie rayonnée est égale à l'énergie qui prend naissance dans les régions centrales.

Quelles sont les modalités possibles de ce « flux » d'énergie? Il en existe essentiellement deux : le transfert radiatif et le transfert convectif (le processus de transfert par conduction, essentiel en physique terrestre, où les températures sont basses, ne joue aucun rôle ici).

le transfert radiatif

Ce mode de transfert de l'énergie est le mieux connu : le véhicule de l'énergie est l'onde électromagnétique elle-même.

Deux facteurs physiques conditionnent essentiellement ce phénomène. En tout premier lieu, l'opacité du gaz constituant l'étoile.

Cette notion d'opacité est très importante en astrophysique. En effet, aux températures qui règnent dans les étoiles, les gaz ne sont pas transparents comme ils le sont aux températures usuelles. En particulier, à l'intérieur du Soleil, un récepteur idéal (et surtout incombustible) ne verrait qu'à quelques centimètres de profondeur, et l'on peut dire que le gaz constituant le Soleil se comporte comme un épais brouillard.

Cette opacité dépend de la température, de la densité et même de la longueur d'onde; elle est donc différente pour les divers rayonnements de l'étoile : ultraviolet, lumière visible, etc. Sa valeur ne varie pas d'une façon simple en fonction de ces divers paramètres. Etant donné qu'elle intervient dans les principales équations qui permettent de calculer ce qui se passe au sein des étoiles, on a établi des tables numériques de cette valeur dans diverses conditions. L'intégration des équations différentielles où elle figure nécessite toujours de longs calculs numériques, qu'on traite le plus souvent à l'aide de machines électroniques.

Le second facteur physique essentiel du transfert radiatif est ce qu'on appelle le « gradient de température » (ou mesure de la variation de la température).

Isolons maintenant, du moins par la pensée, une couche sphérique au sein de l'étoile. Elle reçoit de la chaleur du centre de l'étoile et elle en émet vers l'extérieur.

On dit qu'entre le centre et le bord règne un gradient de température, et c'est lui qui règle le « flux » d'énergie. On peut comparer le rôle du gradient de température dans le flux d'énergie à celui de la « pente » d'un cours d'eau dans son débit.

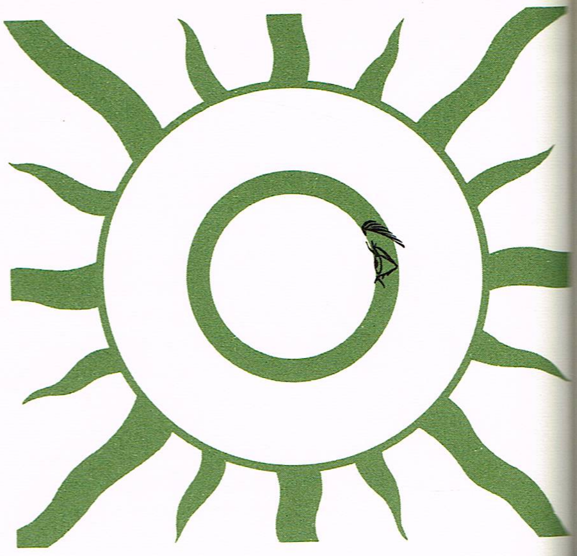
Revenons à un autre récepteur idéal... et incombustible, et supposons-le situé à l'intérieur de l'étoile (ci-contre). S'il « regarde » vers l'intérieur, il verra des régions plus chaudes et il en recevra un rayonnement intense. Inversement, s'il regarde vers le dehors, il verra des régions plus froides. C'est ici qu'intervient l'opacité des gaz.

Si le gaz est transparent, le récepteur verra très loin dans un sens ou dans l'autre. En regardant vers le centre il recevra un rayonnement beaucoup plus intense que s'il se tourne vers l'extérieur. Alors, le flux net d'énergie rayonnée vers l'extérieur sera extrêmement grand. On dit aussi qu'il y aura une grande « anisotropie » du rayonnement.

Inversement, si le gaz est très opaque, l'instrument d'observation se trouvera environné d'une sorte d'atmosphère cotonneuse. Le rayonnement ne lui parviendra plus que des régions immédiatement voisines; le « flux net » sera très faible. Il y aura « isotropie » du rayonnement, et la situation sera très proche de ce qu'elle serait dans une enceinte isotherme à équilibre thermique. C'est-à-dire, finalement, que le rayonnement sera proche de celui du corps noir.

Nous voyons ainsi l'imbrication des deux notions fondamentales d'opacité et de gradient de température. Ces notions qualitatives sont en fait rassemblées en une seule équation différentielle, dite « équation du transfert radiatif », qui est à la base même des études sur l'intérieur des étoiles. En effet, c'est uniquement par voie de rayonnement que se propage l'énergie au sein de toute étoile.

L'œil dans le Soleil.



le transfert convectif

Mais si l'on aborde l'étude des régions centrales, où les températures sont extrêmement élevées, un autre phénomène devient prépondérant : il s'agit du transfert convectif.

L'énergie calorifique peut être transportée au sein des étoiles par de vastes mouvements internes, les courants de convection.

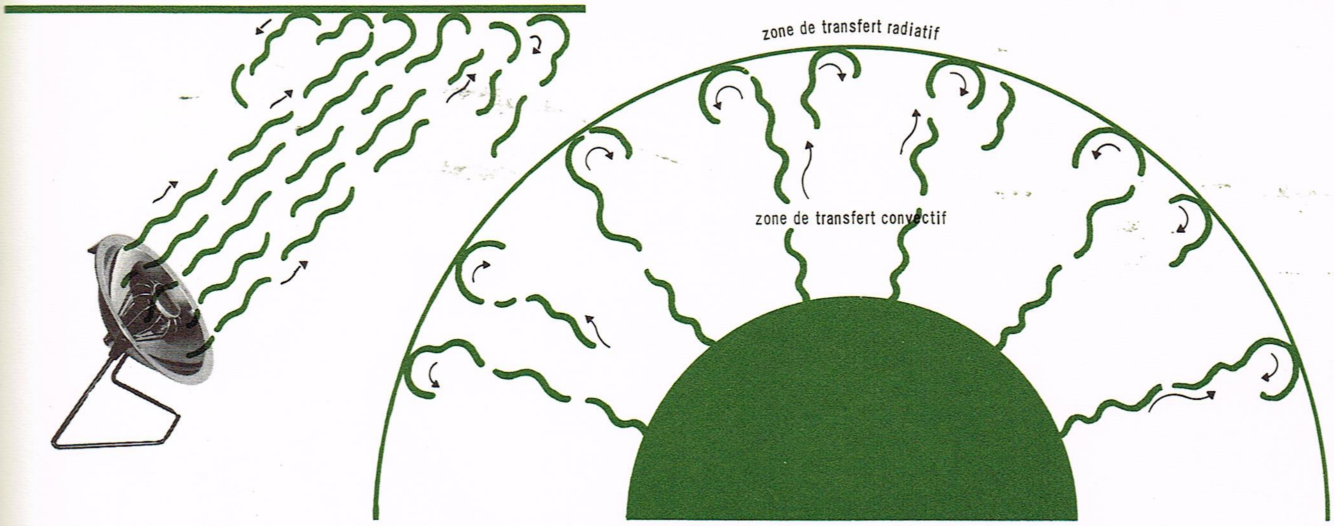
Ces mouvements apparaissent quand le gradient de température est extrêmement grand, c'est-à-dire quand les variations de température entre le centre et les couches externes sont trop élevées pour que ces couches externes gazeuses restent en équilibre les unes au-dessus des autres.

Une étoile dont le centre est beaucoup plus chaud que la surface se comporte comme le font une installation de chauffage central ou un radiateur. De tels mouvements gazeux n'apparaissent cependant pas nécessairement; leur existence dépend non seulement du gradient de température, mais encore du gradient de pression.

Encore une fois, utilisons une image pour faire comprendre cette nouvelle notion.

Imaginons un ballonnet plein d'air chauffé à sa partie inférieure comme les aérostats de nos arrière-grands-parents. Si le ballon s'élève, sa pression interne reste continuellement égale à celle de l'air extérieur avec lequel il est en communication par sa « manche ». Il va rencontrer dans son ascension des couches d'air où la pression atmosphérique est de plus en plus faible. Sa température va diminuer, car tout gaz qui subit une détente (baisse de pres-

volutes d'un radiateur dit parabolique



sion interne) se refroidit. On sait calculer la chute de température engendrée par une diminution donnée de la pression. On appelle cette relation la « loi de la détente adiabatique ». Nous pouvons supposer cette détente assez rapide pour qu'il n'y ait aucun échange de chaleur avec l'air du dehors.

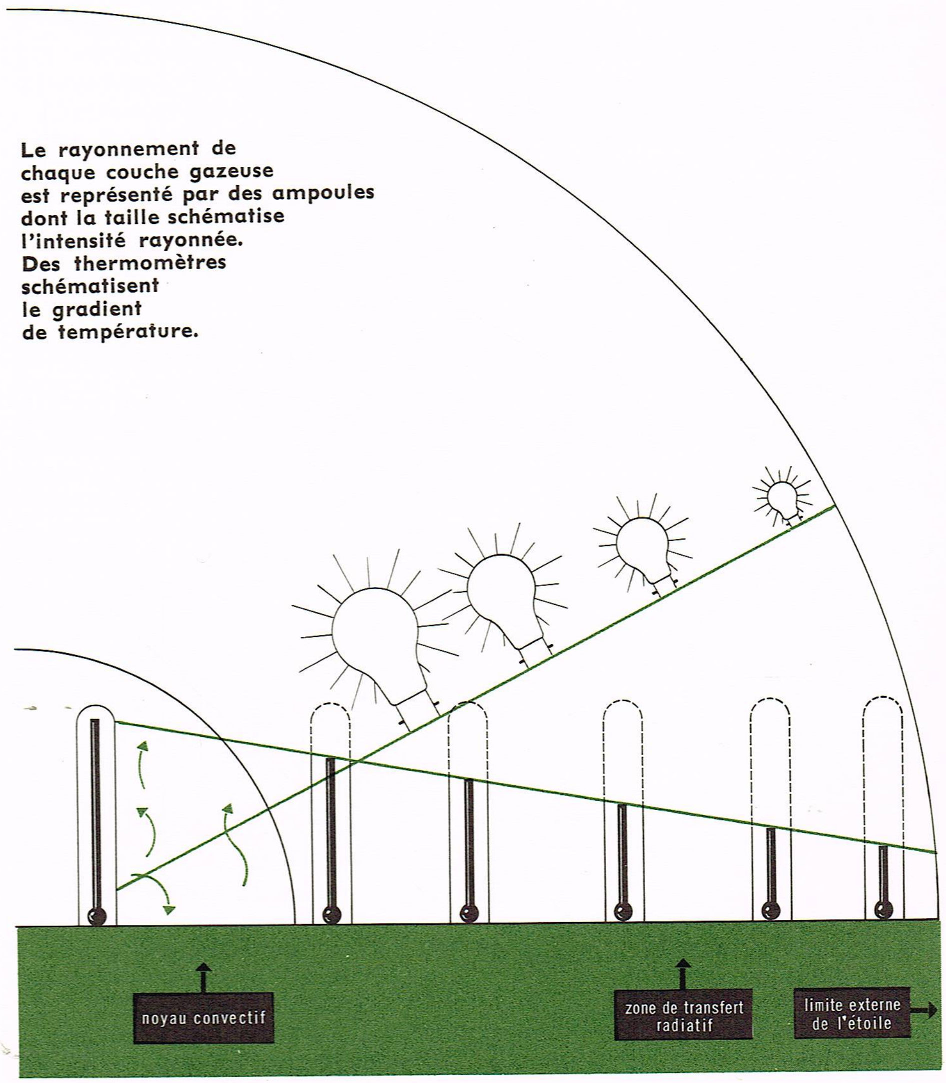
Dans son ascension, le ballonnet peut rencontrer des régions où la température est plus grande que la sienne propre. L'air du ballonnet sera alors relativement plus froid, donc plus lourd que le gaz environnant, et le ballonnet redescendra vers sa position initiale.

Si, au contraire, dans son ascension, le ballonnet rencontre des couches gazeuses plus froides que lui, il se comportera comme un « plus léger que l'air » et continuera à monter. Le cas limite est évidemment celui où, dans l'atmosphère et dans le ballonnet, la pression et la température varient de la même façon. On dit, en pensant précisément à l'analogie avec un ballonnet de gaz isolé, que le gradient de température est le gradient adiabatique dans l'atmosphère considérée.

C'est cette condition-là (condition de l'équilibre convectif) qui règle dans les couches profondes des étoiles l'apparition des courants de convection. Lorsqu'elle n'est pas réalisée, le transport d'énergie du centre vers le bord par les courants de convection est d'autant plus intense qu'il y a un plus fort gradient de température, exactement comme dans le cas du transfert radiatif.

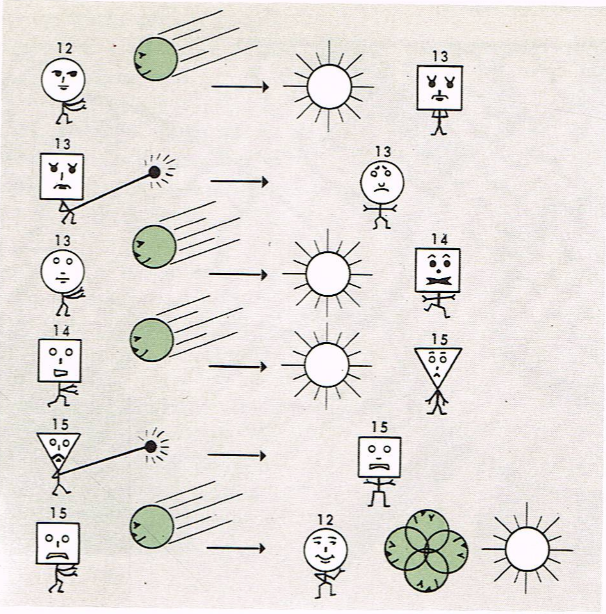
Comme l'a suggéré en premier lieu le grand astrophysicien Eddington, on admet

Le rayonnement de chaque couche gazeuse est représenté par des ampoules dont la taille schématise l'intensité rayonnée. Des thermomètres schématisent le gradient de température.



$$
\begin{array}{lll}
1) & C^{12}_{6} + H^{1}_{1} \longrightarrow N^{13}_{7} + \gamma \\
2) & N^{13}_{7} \longrightarrow C^{13}_{6} + e^{+} \\
3) & C^{13}_{6} + H^{1}_{1} \longrightarrow N^{14}_{7} + \gamma \\
4) & N^{14}_{7} + H^{1}_{1} \longrightarrow O^{15}_{8} + \gamma \\
5) & O^{15}_{8} \longrightarrow N^{15}_{7} + e^{+} \\
6) & N^{15}_{7} + H^{1}_{1} \longrightarrow C^{12}_{6} + He^{4}_{2}
\end{array}
$$

$$\text{soit, au total : } 4\,H^{1}_{1} \longrightarrow He^{4}_{2} + 2e^{+} + 4\gamma$$



Le cycle de Bethe et ses six étapes, dont quatre (équations nos 1, 3, 4 et 6) sont des transmutations pratiquement instantanées et deux (les nos 2 et 5) sont des décompositions radio-actives différées. Dans ce cycle, le carbone, se comportant en quelque sorte comme un catalyseur chimique, se trouve finalement régénéré et prêt à servir de nouveau, tandis que 4 atomes d'hydrogène ont disparu et que 1 atome d'hélium s'est formé.
On retrouve ces équations dans le dessin humoristique fait par l'astronome américaine C. Payne-Gaposchkin.

finalement aujourd'hui que dans la plus grande partie de l'étoile l'énergie est transportée par le rayonnement. Dans les parties centrales, les courants de convection peuvent jouer un rôle important. On désignera souvent ces régions par le nom de « noyau convectif ». Dans les deux cas, le transfert d'énergie est lié à l'existence d'un gradient de température entre le centre et le bord.

l'atome, source de l'énergie stellaire

Pour nous faire une idée de la structure d'une étoile, et plus encore pour comprendre son évolution, il nous faut maintenant aborder une question essentielle : d'où provient l'énergie rayonnée par les étoiles ? C'est une question qui a fait de grands progrès depuis une trentaine d'années. En effet, jusqu'à une époque récente, on ne pouvait concevoir quelle était la formidable réserve d'énergie qui permettait au Soleil de nous dispenser chaleur et lumière depuis des milliards d'années.

On avait bien pensé à une source d'énergie mécanique, la gravitation. Une étoile, pensait-on, en se comprimant, s'échauffe et rayonne. C'est la « chute » des parties extérieures vers le centre qui est la source d'énergie, exactement comme dans les vieilles pendules de nos campagnes, où la chute d'un poids met le mécanisme d'horlogerie en mouvement.

Mais cette source d'énergie se révélait nettement insuffisante, et, à l'heure actuelle, on admet qu'elle intervient seulement au cours d'une période préliminaire, lorsque l'étoile n'est encore qu'une « proto-étoile ». Nous y reviendrons.

Le lecteur a certainement deviné quelle est cette formidable source d'énergie qui fait briller le Soleil et les étoiles, entretenant de plus la vie dans l'univers... C'est l'énergie nucléaire. On sait qu'au début de ce siècle Einstein avait prévu la possibilité de transformer la matière en énergie. La fameuse équation d'Einstein $W = mc^2$ donne le bilan de cette conversion. A l'heure actuelle, on n'a pas encore conçu un mécanisme qui permettrait de convertir en énergie la totalité d'un atome matériel. On ne connaît, on ne sait réaliser que des réactions nucléaires au cours desquelles certains atomes se transmutent en d'autres. (C'est seulement la différence entre la masse totale des premiers et la masse des seconds qui se transforme en énergie.)

Au centre de l'étoile, nous l'avons vu, règnent des températures qui se chiffrent par millions de degrés. Ces conditions sont favorables à des réactions nucléaires qui ne se produisent jamais spontanément sur la Terre, et notamment la fusion de quatre atomes d'hydrogène en un seul atome d'hélium, avec une perte de masse qui se traduit par une libération d'énergie. Divers processus (cycles du carbone, dit « cycle de Bethe », ou cycle proton-proton) sont possibles : tout dépend de la pression et surtout de la température.

Quoi qu'il en soit, le bilan de cette chimie nucléaire est ici simple : « **C'est en transformant dans sa région centrale, toutes les secondes, 564 millions de tonnes d'hydrogène en 560 millions de tonnes d'hélium, que le Soleil produit les 4 millions de tonnes (selon l'équation d'Einstein) de lumière qu'il émet par seconde** ».

la « perforatrice analytique »

Nous venons de dresser un tableau d'ensemble des conditions physiques qui règnent à l'intérieur des étoiles et des divers états de la matière et de l'énergie qui doivent s'y trouver.

Le mathématicien nous dirait que ces conditions se traduisent par des équations différentielles et que, **connaissant approximativement l'état physique de l'étoile à sa surface** (« **conditions aux limites** »), **nous sommes en mesure de calculer, au prix de maintes difficultés, certes, ce qui se passe en dedans.** On peut montrer qu'un tel système d'équations mathématiques n'admet qu'une solution : **la structure d'une étoile est entièrement déterminée par sa masse, sa composition chimique initiale et son âge.**

Ce théorème, dû à Vogt, permet de mieux connaître le centre d'une étoile située à des milliers d'années de lumière que ce qui se passe quelques kilomètres sous nos pieds. Russell à pu en parler comme d'une véritable « perforatrice analytique ».

Comme ces grandeurs sont en rapport avec les autres caractéristiques d'une étoile, en particulier sa luminosité et sa température de surface, il importe de vérifier par des moyens pratiques l'existence de relations entre ces paramètres.

L'une des plus remarquables parmi ces vérifications est possible grâce à l'existence d'une relation masse-luminosité. **Il n'est pas exagéré de dire qu'avec le diagramme de Hertzsprung-Russell, dont nous avons déjà parlé, cette relation constitue l'un des fondements de toutes les théories qui concernent la structure et l'évolution des étoiles.**

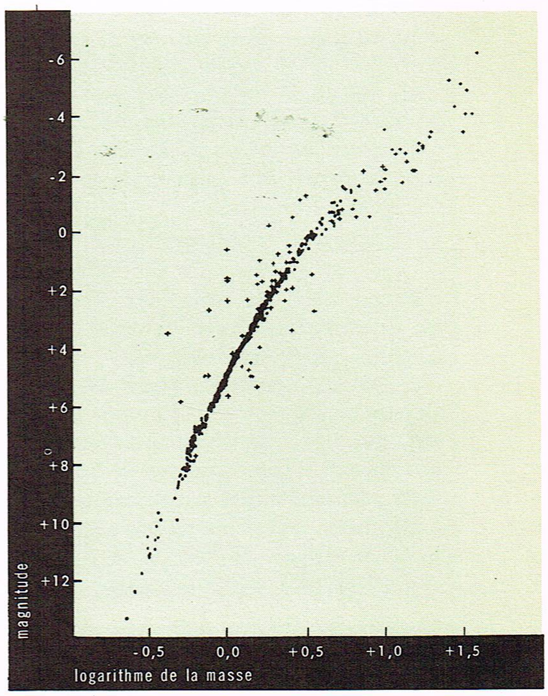
Si nous portons en ordonnées la luminosité (exprimée en magnitude absolue) et en abscisses la masse, nous constatons que la plupart des étoiles dont les masses sont connues se groupent autour d'une ligne traversant le diagramme de gauche à droite.

Certaines étoiles pourtant n'obéissent pas à cette relation. C'est qu'elles sont construites sur un autre modèle, et aussi qu'elles ont probablement des compositions chimiques légèrement différentes. Ce sont des étoiles à un autre stade d'évolution que celles de la « série principale », lesquelles se conforment à la relation masse-luminosité.

Le lecteur se rappelle encore qu'au chapitre précédent nous avons parlé du diagramme de Hertzsprung-Russell. Remarquons ici qu'il fait partie, lui aussi, des grandes vérifications qui contribuent à étayer le principe de Vogt. En effet, pour des étoiles qui sont sensiblement au même stade évolutif, par exemple celles de la « série principale » définie ci-dessus, la masse détermine de façon unique la luminosité et la température superficielle, c'est-à-dire le type spectral. On peut alors s'attendre à trouver une relation entre la luminosité et le type spectral.

On retrouve donc, ici encore, le diagramme de Hertzsprung-Russell, véritable carrefour par lequel passent toutes les branches de l'astrophysique moderne; les travaux qui concernent sa détermination et son explication alimentent l'astronomie moderne depuis cinquante ans.

Le lecteur peut maintenant se rendre compte de la difficile et patiente investigation qui a conduit les chercheurs jusqu'au noyau central des étoiles. Nous avons vu comment on pouvait être amené à estimer valablement la nature chimique, la densité, la température, la pression de sa matière, ainsi que les modalités de transfert de l'énergie dégagée, et jusqu'à la source même de cette énergie.



Relation masse-luminosité. Chaque point représente une étoile.

(D'après Franz.)

les étoiles vivent et meurent aussi

Le ciel étoilé représente pour le profane l'image même de l'immuable. Et pourtant, à y regarder de plus près, cette fixité n'a rien d'absolu.

Nous avons déjà parlé des divers mouvements qui animent les étoiles. Celles-ci ne sont pas immobiles dans l'espace. De même, la fixité de leur structure physique cesse d'être absolue dès que l'on considère des échelles de temps dépassant les quelques milliards d'années qu'on attribue à notre système solaire sous sa forme actuelle.

Quelles sont donc les étapes de la vie d'une étoile? Où prend-elle naissance? Comment meurt-elle? Voici des problèmes qui sont parmi les plus actuels de l'astrophysique, et qu'on eût à peine osé formuler il y a une vingtaine d'années.

Ce sont Gamow et Teller qui, en 1945, donnèrent un premier schéma des étapes évolutives que nous allons décrire. C'était alors œuvre d'intuition plus que de science; les théories de ces auteurs ne furent pas immédiatement suivies par l'ensemble des astrophysiciens. Mais, une dizaine d'années plus tard, des tests solides vinrent étayer ces premières tentatives. Le problème est encore lui-même en pleine évolution. Parmi les savants qui contribuèrent le plus à son développement, nous ne pouvons manquer de citer quelques grands noms : Hoyle, Schwarzschild et Sandage.

Abordons maintenant les phases successives de la vie d'une étoile, telle qu'on l'imagine aujourd'hui. Nous décrirons d'abord la vie et la mort d'un astre, puis nous tenterons — projet autrement ambitieux — de montrer comment on est arrivé à cette conception et quelles en sont les preuves expérimentales.

comment naît une étoile

Si, par une nuit claire, nous regardons la Voie lactée, nous y voyons certaines parties plus obscures, comme si les étoiles avaient boudé ces régions. Il n'en est rien. On sait maintenant que ces espaces nous paraissent dépourvus d'étoiles parce qu'ils sont remplis par un « gaz » absorbant formé d'hydrogène et de poussières : la matière interstellaire, qui occupe ainsi la presque totalité de l'espace galactique.

On a longtemps cherché quels peuvent être les rapports entre cette matière interstellaire et les étoiles.

Les films que l'on prend du Soleil montrent qu'une étoile éjecte d'immenses masses de gaz. Aussi semblait-il logique de penser que la matière interstellaire est un sous-produit de l'activité des étoiles; mais cette attitude a dû être en partie abandonnée; certes, les étoiles rejettent des gaz, mais on admet aujourd'hui que le phénomène inverse existe aussi et qu'il est à l'origine de la formation des étoiles : c'est la *théorie de l' « accrétion »*.

Comment cela se produit-il ?

Une galaxie — la nôtre, par exemple — est un immense disque de matière interstellaire, contenant plusieurs milliards d'étoiles, qui tourne à la façon d'une roue. On sait maintenant qu'à l'origine il n'existait que de la matière interstellaire, mais que celle-ci se condense par endroits, formant dans certaines zones des sortes de « nuages ». Avec le temps, un nuage galactique se coagule en taches, se contracte, devient de plus en plus dense.

Pourquoi ?... Mais simplement à cause des forces de gravitation. Dès qu'un noyau de condensation s'est formé, il apparaît une « force de gravité » analogue à ce qu'est la pesanteur sur la Terre. Alors les gaz voisins « tombent », si l'on peut dire, sur le noyau initial. A ce stade, nous n'avons pas encore une étoile, mais une « proto-étoile » froide. A mesure que la contraction progresse, le gaz s'échauffe. Lorsqu'il atteint une température suffisante, il commence à rayonner. Un point lumineux apparaît au ciel : une étoile est née.

devenue adulte, l'étoile se stabilise

La contraction se poursuit toujours, et la température continue à monter jusqu'à ce qu'elle soit suffisante pour produire des réactions nucléaires au centre de l'étoile.

On admet que la température centrale doit dépasser 10 millions de degrés pour que ces processus prennent naissance.

Nous avons alors une étoile à l'âge adulte, dont la structure répond à celle que nous avons décrite dans le précédent chapitre : au centre, la fournaise nucléaire; autour d'elle, une zone de « transfert convectif »; et, à la périphérie, une zone de « transfert radiatif ».

En termes de classification spectrale, l'étoile, à ce stade, est typiquement une étoile de la séquence principale du diagramme de Hertzsprung-Russell. C'est le cas de notre Soleil, par exemple. Et c'est cet état, que nous pourrions qualifier de « normal », que l'étoile gardera la plus grande partie de son existence.

Comment expliquer la grande variété des étoiles de la séquence principale, depuis les étoiles rouges, presque froides, jusqu'aux étoiles bleues, dont la température approche, en surface, de 50 000 °C?

Cette diversité est causée par les différences de masses entre les nuages interstellaires initiaux. La figure ci-dessous, due à Hoyle, montre comment les étoiles de différentes masses se placent en des points différents du diagramme de Hertzsprung-Russell.

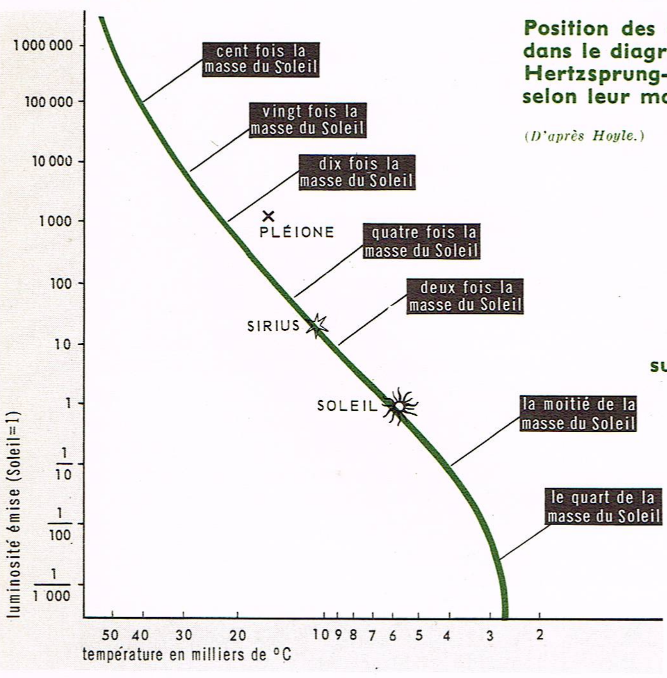
un astre sur son déclin

L'hydrogène s'épuise au centre; sa transformation progressive en hélium par le moyen des réactions thermonucléaires aboutit à la formation d'un noyau central inerte.

Ce qui se passe alors ressemble à un incendie dans une forêt ou, mieux, dans une savane : si le foyer initial est bien délimité, il arrive un moment où tout est consumé au centre. Le feu continue à progresser sur la périphérie d'une région centrale dénudée, et celle-ci s'agrandit de plus en plus. De même, on admet que les réactions nucléaires se produisent, à ce stade, dans une fine pellicule entourant la région centrale d'hélium inerte et qui se déplace peu à peu vers l'extérieur.

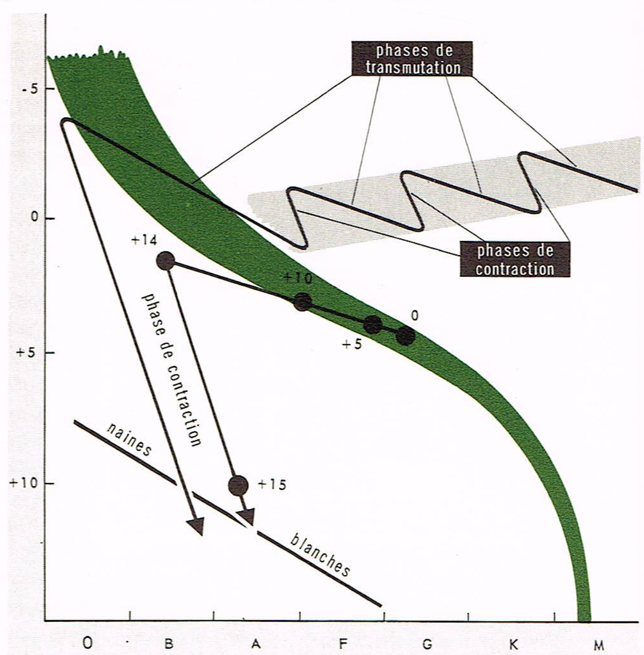
Pour les étoiles plus grosses, le noyau d'hélium se contracte sous l'effet de la pression des couches extérieures. Par le jeu combiné des fortes pressions et des températures élevées, on peut concevoir que l'hélium lui-même entre en réaction nucléaire en donnant naissance à des atomes plus lourds, tels que l'oxygène. Nous avons alors une nouvelle source d'énergie au centre de l'étoile, et il se produit une expansion considérable des enveloppes extérieures; l'étoile devient géante.

Quand le Soleil arrivera à ce stade, son diamètre sera plusieurs centaines de fois supérieur à ce qu'il est actuellement, c'est-à-dire que cet astre aura englouti toutes ses planètes, y compris la Terre.



Position des étoiles dans le diagramme de Hertzsprung-Russell selon leur masse

(D'après Hoyle.)



Évolution probable de deux étoiles sur un diagramme de Hertzsprung-Russell. L'une naine, le Soleil, figurée de cinq en cinq milliards d'années; l'autre, géante, à travers des phases de contraction et de transmutation successives.

Phot. Mayall et Babcock, télescope de l'observ. Lick.

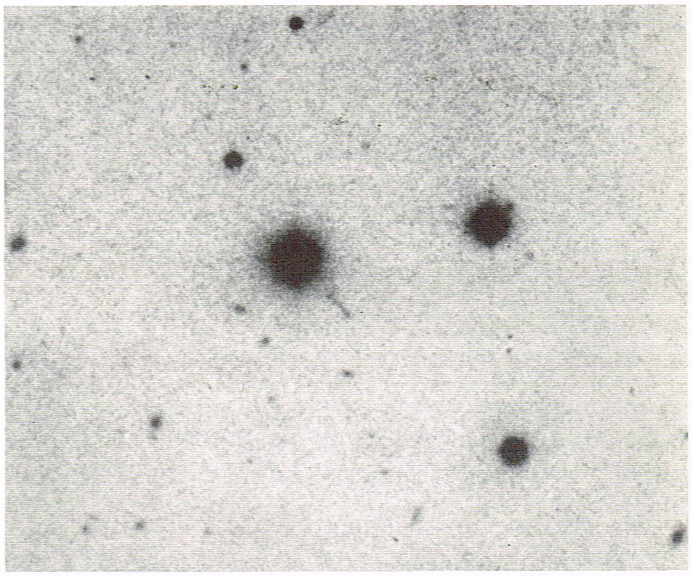
les quasars

Après la découverte des radiogalaxies, les astronomes furent frappés par un fait curieux : les étoiles, si brillantes en optique, ne correspondaient jamais à une radiosource, sauf dans deux cas. Leur attention se porta donc sur ces étoiles exceptionnelles; des photographies détaillées furent prises au Mont-Palomar, et autour de l'une d'entre elles on découvrit une petite nébulosité que l'on attribua à du gaz éjecté par l'étoile. Des spectres furent pris et là, stupeur, on s'aperçut que les raies des éléments chimiques leur donnant naissance avaient des longueurs d'onde 15 p. 100 plus grandes que la normale.

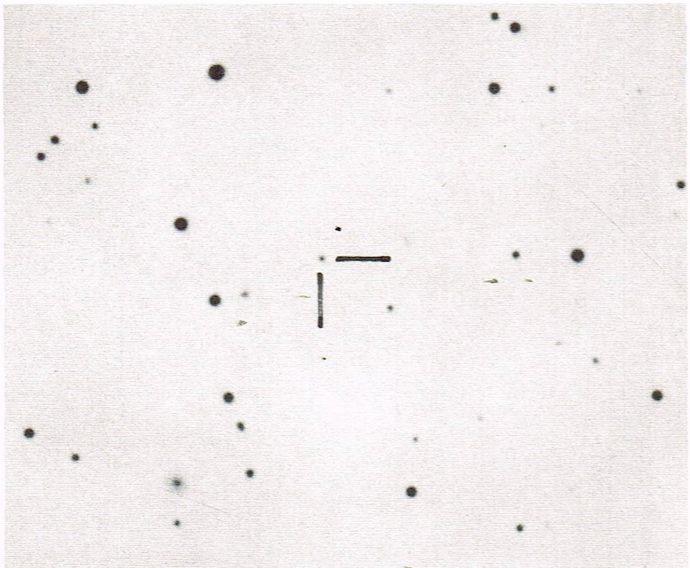
Après avoir analysé les causes possibles de ces décalages spectraux, les astronomes américains furent contraints d'admettre que ces décalages étaient dus à une grande vitesse d'éloignement de ces astres : 45 000 km/s, comme pour les galaxies soumises à l'expansion de l'univers. Une analyse serrée des diverses hypothèses conduisait à admettre, en plus, qu'effectivement, comme les galaxies, ces objets semblaient s'éloigner de nous à grande vitesse parce qu'ils étaient à grande distance. D'ailleurs, par la suite, un examen attentif des dernières observations radio-astronomiques, la prise de nouveaux spectres conduisirent à la découverte de nouvelles « étoiles », dont certaines semblaient s'éloigner de nous à 600 000 km/s, soit deux fois la vitesse de la lumière, si on se sert de la formule classique de Doppler pour évaluer les vitesses à partir des décalages spectraux.

Deux conclusions immédiates découlent de ces observations : d'abord, une conclusion intéressant la nature physique de ces astres. Nous paraissant brillants comme des étoiles et étant, en réalité, un million de fois plus lointains qu'elles, ils sont donc un million de millions de fois plus brillants intrinsèquement que des étoiles. Ce chiffre est formidable si l'on songe qu'une galaxie comme la nôtre est déjà brillante comme 10 000 millions de soleils; chacune de ces « étoiles » est donc aussi brillante que 100 galaxies. Ce sont de loin les astres les plus brillants intrinsèquement que l'on connaisse dans l'univers. D'un autre côté, leur aspect quasi ponctuel sur les photographies, leur aspect « quasi stellaire », nous montre que l'énergie énorme rayonnée par ces astres est, de plus, rayonnée à partir d'un volume extrêmement faible, bien plus petit qu'une galaxie. Et, en radio, la situation est la même.

Connaissant la distance et le rayonnement de ces astres, on peut calculer que si le rayonnement d'une telle « radiosource quasi stellaire » se maintient tel qu'il est pendant 100 000 ans, ce qui semble un minimum, l'énergie libérée dans ce processus est de l'ordre de 10^{61} ergs. Cette énergie correspond à la masse de 10 millions de soleils; c'est-à-dire qu'un « quasar » (contraction de « quasi stellar ») libère l'énergie que l'on obtiendrait en annihilant entièrement la masse de 10 millions de soleils. Rappelons que, dans une réaction thermonucléaire, seuls quelques millièmes de la masse sont transformés en énergie. Quel phénomène explosif a pu donner naissance à de tels dégagements d'énergie, voilà la question de physique fondamentale actuellement posée par les quasars.



Le premier quasar découvert : 3C 273. C'est aussi le plus proche connu; il ne nous fuit qu'à 45 000 km/s. On remarque le petit jet semblant s'échapper du quasar lui-même; ce jet n'est pas formé de gaz comme on le croyait au début, mais de radiation synchrotron.



Le plus lointain quasar connu : 3C 9. La petite « étoile » au centre est, en réalité, aussi brillante que 1 000 milliards de soleils et les rayons qui ont impressionné cette plaque photographique sont partis à une époque où la Terre n'existait pas encore.

Seconde conséquence importante : l'étude de l'expansion de l'univers, qui plafonnait au voisinage de vitesses de 150 000 km/s, fait un nouveau bond en avant. En effet, les quasars étant intrinsèquement cent fois plus lumineux qu'une galaxie, on doit pouvoir les observer de dix fois plus loin. Or, explorer l'univers dix fois plus loin, c'est en explorer un volume mille fois plus grand, ce qui est fondamental pour étudier sa structure spatiale.

De plus, on peut, en observant les quasars, remonter dans le passé de l'univers bien plus loin qu'avec les galaxies. Avec les instruments actuels comme le télescope du Mont-Palomar ou le radiotélescope de Nançay, on peut déjà remonter dans le passé jusqu'à s'approcher d'un milliard d'années seulement du début de l'expansion, c'est-à-dire étudier 90 p. 100 de l'histoire du monde. Des effets évolutifs importants ont été mis en évidence : ainsi, il semble que les quasars étaient beaucoup plus fréquents au début que maintenant; ils nous apparaissent un peu comme les étincelles d'un gigantesque feu d'artifice initial.

Enfin, dernier point fondamental : on sait que les lois de la relativité interdisent à tout corps de se déplacer plus vite que la vitesse de la lumière, soit 300 000 km/s. Donc, observer des astres *semblant* s'éloigner de nous à 600 000 km/s présente un intérêt prodigieux et permet d'étudier des effets relativistes violents intéressant l'univers dans sa totalité.

Ces trois possibilités importantes apportées par les quasars :
— exploration profonde de l'espace,
— remontée dans le passé lointain,
— effets relativistes extrêmes,
nous font entrer de plain-pied dans l'étude, par l'observation, de la cosmologie.

Clichés négatifs Mont Palomar.

l'inéluctable vieillesse

Cette phase est la moins bien connue. Dans certains cas, on peut prévoir une évolution explosive. L'étoile se brisera en quelques minutes, répandant ses fragments dans l'univers. Mais on peut également prévoir une fin paisible, caractérisée par une diminution générale du volume et de l'éclat. Paradoxalement, l'étoile devient plus bleue et finit ses jours sous une forme encore mal connue, celle de « naine blanche ». Ces étoiles sont extrêmement petites — de la taille de Jupiter —, mais leur densité est des milliers de fois plus grande que celle de cette planète (Kuiper a découvert en 1934 une étoile qui est de moitié plus petite que la Terre, très chaude, et tellement dense qu'un centimètre cube y pèse 36 tonnes). Elles ont un faible éclat intrinsèque, ce qui rend leur étude difficile. En raison des densités énormes qui y règnent, on est amené à penser qu'il existe dans ces étoiles des conditions physiques inconnues sur notre Terre. A ce stade, l'évolution est extrêmement lente, et ses étapes sont encore en grande partie ignorées.

vérifier la théorie de l'évolution ?

Il importe de se demander si ce que nous venons d'exposer est une pure vue de l'esprit ou s'il en existe quelques preuves tangibles.

Nous sommes partis de l'hypothèse fondamentale que voici : les étoiles sont issues de condensations qui se produisent dans les nuages épais de gaz interstellaire.

A un stade plus avancé, voici le système solaire. L'astronomie nous montre, par l'analyse des vitesses de rotation, qu'à l'origine le noyau solaire englobait toutes les planètes, jusqu'aux plus lointaines.

Mais ce n'est là encore que présomptions. Des preuves beaucoup plus sûres et plus précises sont fournies par l'étude systématique des amas d'étoiles et de leur diagramme de Hertzsprung-Russell.

L'importance de l'étude des amas et de leur diagramme de Hertzsprung-Russell avait été prévue par Baade, un des plus grands noms de l'astronomie contemporaine. Les travaux longs et minutieux qui s'y rapportent sont l'œuvre actuelle de Sandage.

Et, tout d'abord, qu'est-ce qu'un amas? On en distingue deux grands types, qui diffèrent par leur morphologie, leur taille, leur stabilité, leur état d'évolution.

Les *amas galactiques*, ou *amas ouverts*, contiennent quelques centaines d'étoiles. Elles sont groupées à la même distance du Soleil et leurs mouvements ne sont pas indépendants. Ces amas sont localisés dans les bras spiraux de la Galaxie et leurs distances peuvent être calculées avec précision si l'on connaît leur mouvement propre. On peut également déterminer ces distances par la méthode des parallaxes spectroscopiques.

Les *amas globulaires*, qui représentent le second type, ont un aspect caractéristique bien connu de tout « astrophile »; on les reconnaît à leur forme sphérique, avec une très grosse condensation au centre et un nombre d'étoiles important : des centaines de mille.

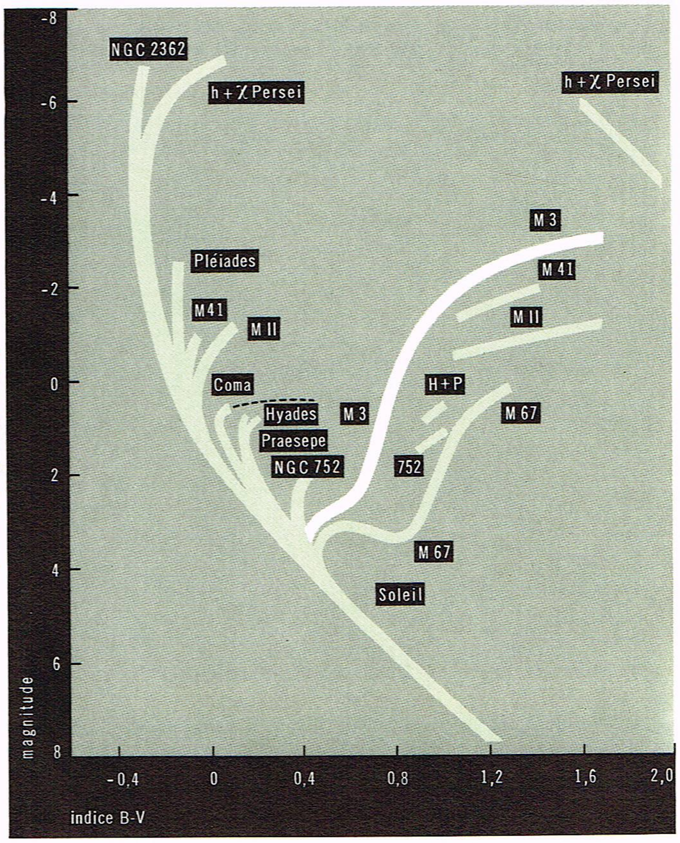
Il est temps de revenir au sujet de nos préoccupations : on pense, en effet, que les étoiles de tels amas ne diffèrent les unes des autres que par un paramètre de base : la masse. La composition chimique et l'âge de toutes ces étoiles doivent être sensiblement les mêmes, car il est logique de supposer qu'elles ont une origine commune. Néanmoins, la diversité de leurs masses entraîne une diversité d'états évolutifs : les étoiles d'un amas ne sont nullement identiques.

Il est, à ce point de vue, particulièrement suggestif de construire le diagramme de Hertzsprung-Russell d'un amas et de le comparer avec un diagramme d'évolution (voir pp. 50 et 58).

L'analogie frappante qu'ils présentent constitue l'une des meilleures preuves de la théorie actuelle de l'évolution des étoiles.

Le diagramme de l'amas montre les diverses phases de cette évolution, depuis l'état initial jusqu'au stade de géante. Dans le cas des amas, on peut dire en quelque sorte que le diagramme de Hertzsprung-Russell est une « photographie instantanée » d'un diagramme de Hertzsprung-Russell dépendant du temps. Comme les étoiles d'un amas se sont formées approximativement à la même époque, et ont sans doute la même composition chimique, les différences entre elles sont d'origine évolutive et dues aux inégalités de masse.

Cela veut dire qu'un diagramme de Hertzsprung-Russell complet serait formé de la superposition de diagrammes partiels correspondant chacun à une date de naissance. Effectivement, en superpo-



Superposition des diagrammes de Hertzsprung-Russell pour plusieurs amas galactiques et pour un amas globulaire (M3). L'indice de couleur B-V est lié au type spectral par une relation simple.

(D'après A. Sandage.)



Les Pléiades et leurs nébulosités.

sant les diagrammes partiels représentant plusieurs amas, on arrive à reconstituer le diagramme de Hertzsprung-Russell tel que nous l'avons décrit dans le paragraphe de la classification des étoiles.

La comparaison des diagrammes évolutifs théoriques et des diagrammes effectivement établis pour les amas d'étoiles ne constitue pas seulement une preuve de la théorie de l'évolution. Elle permet aussi une estimation de l'âge des étoiles.

Sans entrer dans les détails techniques, indiquons quelques résultats : l'amas des Pléiades a environ 100 millions d'années, et celui des Hyades, un milliard.

Certains amas très jeunes, avec des étoiles très brillantes, semblent âgés de moins de dix millions d'années.

Quant aux plus âgés, comme Messier 3, ils ont à peu près de six à neuf milliards d'années, guère plus que les quatre milliards attribués à notre globe et au système solaire.

Le lecteur doit bien comprendre qu'il ne s'agit pas de calculer l'âge d'une étoile à l'aide de documents précis, datés. Il s'agit de faire coïncider les courbes théoriques, établies par le calcul, avec des courbes résultant de l'accumulation de points expérimentaux. On fait, en quelque sorte, glisser la seconde courbe jusqu'à ce qu'elle se superpose à la première.

Mais la précision n'est pas très grande et, de plus, la forme des courbes théoriques est très sensible à la composition chimique supposée de l'étoile, qui ne nous est pas connue exactement.

On ne s'étonnera donc pas de voir tel auteur, parmi les plus éminents, annoncer pour tel amas d'étoiles un âge double de celui qui vient d'être annoncé par tel autre, non moins digne de confiance. Un phénomène identique se produit pour l'estimation des distances les plus lointaines; la presse a fait grand bruit, voici quelques années, au sujet d'une découverte de Baade qui doublait les dimensions de l'univers. Aux extrêmes confins de la spéculation astronomique ces tâtonnements sont chose absolument normale.

Phot. E. E. Barnard, observatoire Yerkes.

un Univers en perpétuelle expansion

Après nous être penchés sur la vie et la mort d'une étoile, posons-nous deux autres questions qui viennent tout naturellement à l'esprit : *D'où vient l'univers? Où va-t-il?*

Ces problèmes font l'objet d'une branche scientifique généralement désignée sous le nom de « cosmogonie »; les questions de l'étendue de l'univers dans l'espace et de sa durée dans le temps y jouent un rôle essentiel. L'association de ces deux mots, espace et temps, évoque aussitôt la théorie de la relativité; et, effectivement, c'est seulement après le grand essor de cette théorie — et plus particulièrement de celle de la relativité générale, il y a un quart de siècle — que la cosmologie est entrée dans son ère scientifique.

C'est à peu près à cette époque que se place une découverte expérimentale de la plus haute importance, que nous allons exposer maintenant. Il s'agit du décalage en direction du rouge de la lumière émise par les galaxies lointaines. En 1920, Slipher annonce ce phénomène, dont les répercussions sur la cosmologie sont considérables : **toutes les nébuleuses lointaines présentent dans leur spectre un important décalage des raies, d'émission ou d'absorption, vers le rouge, et ce décalage paraît être d'autant plus grand que les nébuleuses sont plus éloignées de nous.**

Or, ainsi que nous l'avons montré précédemment, le décalage des raies vers le rouge peut s'interpréter, en vertu de l'effet Doppler, comme le signe d'un éloignement plus ou moins rapide de la source lumineuse par rapport à l'observateur. On parle, dans ce cas, de « vitesse de récession ».

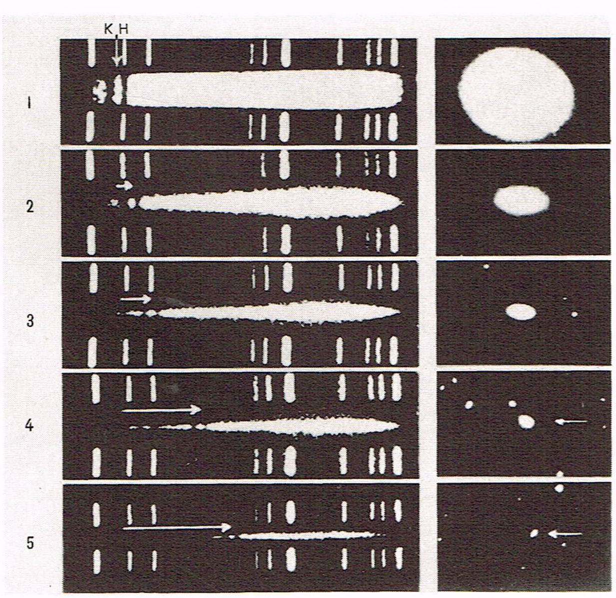
En 1929, à l'observatoire du Mont-Wilson, Hubble montre que les nébuleuses peuvent être « résolues » en systèmes d'étoiles. La présence d'étoiles variables rend possible l'évaluation de leur distance. Toutes ces nébuleuses ont une propriété extrêmement remarquable : la vitesse avec laquelle elles s'éloignent de nous s'accroît proportionnellement à leur distance. **En d'autres termes, il existe une relation linéaire entre les vitesses de ces galaxies et la distance qui les sépare de nous!**

Allons-nous redevenir, du même coup, le centre de l'univers? Une telle découverte aurait pu y conduire, si elle avait été faite plusieurs siècles plus tôt... mais, au XXe siècle, l'interprétation est à la fois plus modeste et plus grandiose : Hubble émet l'hypothèse que l'univers, considéré comme un tout, est en expansion continuelle.

Une longue période de travail expérimental a suivi cette découverte. De 1928 à 1936, Humason photographia le spectre de milliers de nébuleuses; mais le plus délicat était de déterminer leur éloignement. De grands progrès furent faits dans cette voie à partir de 1951, lorsqu'on put utiliser le télescope géant du Mont-Palomar, dont ce fut l'une des plus importantes fonctions. Le phénomène n'est nullement faible et difficile à mettre en évidence; il s'agit, au contraire, d'un effet considérable amenant les raies d'absorption de la région bleue à apparaître dans le rouge, c'est-à-dire à l'autre extrémité du spectre. Le phénomène ne commence, cependant, à se manifester qu'à partir d'une distance déjà considérable. Les « vitesses de récession » mesurées sont d'un ordre de grandeur bien supérieur à celui des vitesses que l'on constate chez les étoiles de notre Galaxie : elles vont de 1 200 à 40 000 et même 60 000 km/s, c'est-à-dire qu'elles atteignent jusqu'au cinquième de la vitesse de la lumière.

On s'est longtemps demandé si ce décalage du spectre devait vraiment s'interpréter comme un effet Doppler, et s'expliquer par une vitesse de récession réelle.

Or, le phénomène présente bien toutes les caractéristiques d'un effet Doppler, il ne peut pas être expliqué par la présence d'une matière absorbante intersidérale, ni par aucun autre phénomène connu : de tous les effets d'optique susceptibles de jouer dans ces conditions, aucun n'est



La relation entre la distance et la vitesse radiale des nébuleuses extragalactiques. Le déplacement des raies H et K de leurs spectres *(à gauche)* est d'autant plus grand que la nébuleuse est plus lointaine.

1. NGC 221 (M 32) d'Andromède : vitesse radiale = — 200 km/s; distance = 750 000 années de lumière;

2. NGC 4473 dans la Vierge : vitesse = + 2 250 km/s; distance = 7 millions d'années de lumière = 2,5 mégaparsecs;

3. NGC 379 des Poissons : vitesse = + 5 500 km/s; distance = 23 millions d'années de lumière = 7 mégaparsecs;

4. Nébuleuse dans l'amas N1 de la Grande Ourse : vitesse = 15 500 km/s; distance = 85 millions d'années de lumière = 26 mégaparsecs;

5. Nébuleuse dans l'amas des Gémeaux : vitesse = + 23 000 km/s; distance = 135 millions d'années de lumière = 42 mégaparsecs.

La faible vitesse apparente d'approche de M 32 résulte de la rotation du Soleil dans la Galaxie. (D'après Hubble et Humason, observatoire du Mont-Wilson, 1935.)

Doc. Observatoire du Mont-Wilson.

capable de modifier la longueur d'onde d'une radiation.

On conçoit que la découverte d'une expansion continue de l'univers, d'autant plus rapide que les régions considérées sont plus éloignées de nous, ait eu un retentissement considérable dans la cosmologie. L'hypothèse que nous venons d'exposer est la plus généralement admise à l'heure actuelle.

l'Univers est-il né de l'atome primitif de G. Lemaître ?

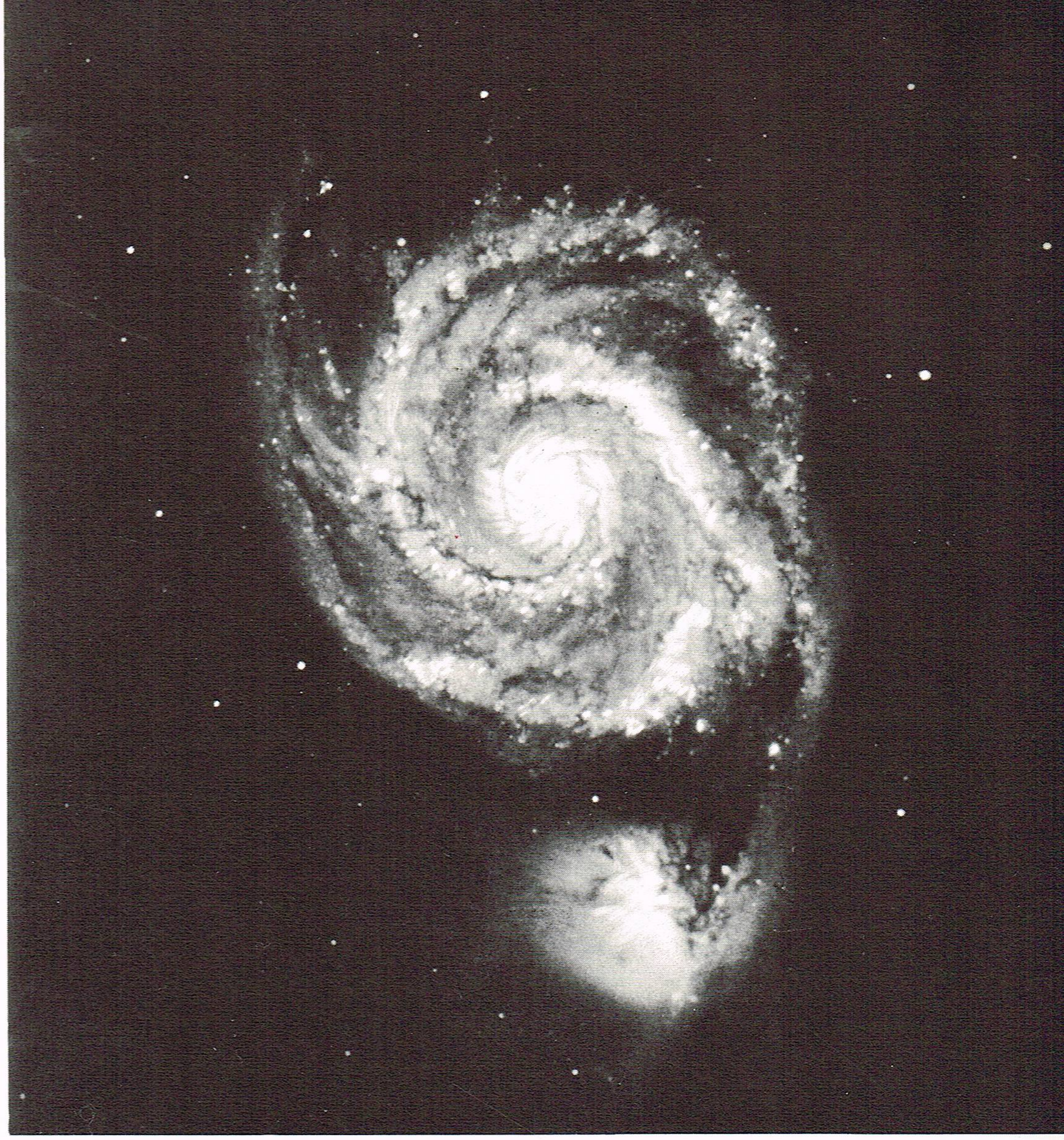
Remontons à présent dans le temps. Selon la théorie de l'expansion de l'univers, il y a des millions de siècles l'univers a dû être beaucoup plus dense qu'il ne l'est aujourd'hui. **La relation entre distance et vitesse nous conduit à penser qu'en un temps reculé toutes les galaxies devaient être concentrées dans une région étroite de l'espace, sorte de noyau initial du monde.** Ce temps reculé, il est facile de le préciser : on le connaît sous le nom de *constante de Hubble*. C'est une date qui se situe à environ dix milliards d'années en arrière. A cette époque, tous les amas de galaxies auraient été serrés les uns contre les autres, donnant à la matière ainsi agglomérée une densité extrêmement élevée.

A mesure que l'expansion se poursuivait, la densité diminuait régulièrement jusqu'à la très faible valeur actuelle. Cette théorie fut tout d'abord développée par un astronome belge, le chanoine Georges Lemaître (1).

Les idées que nous venons d'esquisser sont communes à toutes les théories cosmologiques actuelles. Mais, à partir de ces conceptions premières, on voit se dessiner deux grands courants.

Le premier, avec comme promoteur le grand savant et écrivain scientifique Gamow, développe la théorie d'un univers dont l'évolution se poursuit indéfiniment dans le même sens. L'autre courant, avec comme chef de file l'astrophysicien Hoyle (déjà cité à propos de l'évolution des étoiles), tend à envisager un univers permanent.

(1) Ce dernier fut d'abord conduit à évaluer l'âge de l'univers à deux milliards d'années à une époque où la naissance de la Terre était considérée comme antérieure à quatre milliards d'années. C'était fâcheux, et il a fallu attendre que des réévaluations successives des distances — donc de la constante de Hubble — eussent porté l'âge de l'univers à plus de six milliards d'années pour que fût résolu ce curieux paradoxe.



Nébuleuse spirale M 51 des Chiens de chasse (15 mai 1926) photographiée avec le télescope de 2,54 m de l'observatoire du Mont-Wilson.

Résumons brièvement les traits marquants de ces deux tendances.

Pour Gamow, qui a repris et précisé l'idée de Lemaître (atome primitif), l'univers a commencé par une explosion à partir d'un état initial super-dense. L'univers est clos, mais en expansion continuelle. Cette expansion se poursuivra de plus en plus lentement, mais indéfiniment. Dans le noyau super-dense qui explosa il y a cinq ou sept milliards d'années, toute la matière de l'univers s'est formée. L'élément matériel originel est le neutron. **Gamow admet qu'en quelques minutes, au début de l'univers, fut créée toute la matière qui constitue étoiles, planètes et galaxies.**

un Univers stable, mais où de la matière se crée sans cesse ?

Hoyle et Bondi ont proposé une théorie selon laquelle, au contraire, l'univers présente un caractère de « permanence », c'est-à-dire que ses grands traits, et notamment sa densité, ne se modifient pas au cours du temps. On ne peut lui assigner ni commencement ni fin.

On ne peut évidemment concilier une densité constante avec un rayon, donc un volume croissant, qu'en admettant que la quantité de matière augmente, ce qui conduit tout naturellement à l'hypothèse

d'une « création continue de matière », en opposition avec la thèse précédente. La matière serait toujours créée à l'état d'hydrogène. Cet élément, par contraction gravitationnelle, formerait des protoétoiles, puis des étoiles où les éléments lourds seraient produits à la suite de réactions nucléaires; enfin, lorsqu'une étoile aurait terminé son existence, son explosion répandrait les éléments dans l'espace.

Théorie séduisante, mais qui se heurte à de si graves difficultés qu'elle a dû être abandonnée, au moins sous cette forme. En vertu du principe de conservation des charges « baryoniques » auquel la physique moderne aurait peine à renoncer sans raison grave, il ne peut y avoir création de matière que s'il y a création d'une « quantité » égale d'antimatière (v. tome II, chap. *Le monde agité des particules élémentaires*); et, bien que ce sujet soit encore débattu, les interactions de matière et d'antimatière devraient produire d'intenses radiations, que l'on n'observe pas.

Il y a plus grave encore : selon cette théorie, l'état chimique de la matière d'une galaxie serait indépendant de son âge (la création d'éléments légers compensant la formation d'éléments lourds), ce qui semble contraire à l'observation.

de ces belles théories quelle est la bonne ?

La théorie évolutionnaire implique que l'expansion, bien que continuelle, diminue progressivement. Or, nous avons vu que le grand télescope de Hale explore un temps de deux milliards d'années. Si la théorie de l'explosion initiale est correcte, l'expansion devait être plus rapide à cette époque qu'elle ne l'est aujourd'hui. Ainsi, nous voyons que le décalage des spectres vers le rouge est un élément de base de la cosmologie moderne. On peut espérer maintenant que les mesures de ce décalage, faites au Mont-Palomar, rendront bientôt possible un choix définitif entre les deux thèses; on saura alors si l'univers est en expansion continue ou si, au contraire, nous devons adopter l'image d'un univers en équilibre permanent, sinon sous la forme proposée par Hoyle, du moins sous une autre qui serait encore à trouver.

Il ne nous appartient pas encore de choisir entre ces deux conceptions cosmogoniques. Ce ne sont, bien entendu, que des hypothèses, auxquelles d'autres succéderont nécessairement. Nul ne peut deviner ce que nos descendants penseront, dans vingt ou cent mille ans (ni même dans dix ans), de ce vaste problème.

L'homme voit des galaxies dans un passé cosmique qu'il estime à deux milliards d'années; s'il parvenait à assister de cette manière aux commencements du monde dans ces lointains espaces, il pourrait s'en faire une idée plus assurée. Mais cette supposition fantastique ne risque guère de se réaliser; il faudrait tripler la taille actuelle des télescopes, ce qui n'est guère possible techniquement. D'où provient l'énergie fabuleuse qui anime la matière de l'univers? Son mouvement a-t-il commencé? Finira-t-il? Et quelle signification peut-on lui attribuer? Il faut convenir que la science aborde ici à des rivages dangereux, aux confins de la métaphysique.

Ajoutons, pour terminer, que l'idée d'une création continue de matière a incité certains savants à supposer l'existence d'univers « parallèles » d'antimatière, totalement imperceptibles pour nous. Mais que se passerait-il à la frontière de ces deux mondes?

l'histoire de l'Univers et celle de la Terre

La cosmogonie est l'ensemble des doctrines qui tentent de savoir et d'expliquer comment est né l'Univers, et comment il a évolué et évoluera encore, tandis qu'on réserve le terme de « cosmologie » à l'étude du monde tel qu'il est aujourd'hui.

Longtemps, l'origine du système solaire a été le principal objet des cosmogonies; mais, aux temps modernes, les frontières de l'univers connu n'ayant cessé de reculer, on s'est attaché de plus en plus à l'histoire des étoiles et des galaxies. La formation du Soleil et de ses planètes n'est plus qu'un épisode dans l'évolution de la matière universelle, et cet épisode est loin d'être unique en son genre : il a dû se reproduire des milliards de fois...

« La découverte de la formation permanente de nouvelles étoiles nous amène à rayer radicalement de notre esprit l'idée de l'origine simultanée de toutes les étoiles », a pu écrire E. Schatzmann. De même, au sein du système que nous habitons, on a renoncé (depuis longtemps d'ailleurs) à penser que le Soleil, la Terre, la Lune et les planètes se sont tous formés au même moment. Déjà au VIe siècle avant J.-C. on supposa que cette formation a été progressive.

La science cherche à se représenter, avec une précision et des chances de probabilité toujours plus grandes, la genèse de l'univers. Elle fait appel pour cela à un nombre important de disciplines : la cinétique, l'optique, la spectrographie, la thermodynamique, les théories de la relativité, la physique nucléaire.

Nous nous limiterons dans ce chapitre à un résumé succinct des principales théories cosmogoniques relatives au système solaire. En effet, les paragraphes précédents ont suffisamment montré comment on se représente aujourd'hui la naissance, la vie et la mort des étoiles et des nébuleuses.

Jetons d'abord un regard sur les idées cosmogoniques anciennes. Nombreuses sont les légendes qui ont pour thème la genèse de l'Univers, du Soleil et des astres. Les symboles qu'elles recouvrent, notamment dans l'Inde antique et chez Moïse, sont parfois très profonds, mais n'appartiennent pas au domaine de l'astronomie.

C'est seulement en Grèce, au VIe siècle avant J.-C., que sont apparues des théories dégagées de toute mythologie. On est saisi d'étonnement et d'admiration

quand on s'aperçoit que ces philosophes helléniques — eux-mêmes interprètes d'une sagesse plus ancienne, cachée dans les sanctuaires — parlent tout naturellement de la nébuleuse primordiale et de la condensation du Soleil, de la Terre et des planètes...

Plus tard, les Latins Lucrèce et Ovide exposèrent des cosmogonies apparentées à celles que nous venons d'esquisser. Mais, entre-temps, en Grèce, Aristote avait fait prévaloir des idées bien différentes, qui, on le sait, transmises aux penseurs du Moyen Age, s'imposèrent avec la force d'un dogme et enrayèrent pour mille ans tout progrès : pour lui, les astres étaient immuables et existaient de toute éternité.

Lors de la Renaissance, la révolution copernicienne allait fournir aux philosophes un tableau cohérent de l'univers.

Pour Descartes, le Soleil et les planètes sont issus d'une sorte de chaos de particules animées d'un mouvement tourbillonnaire et qu'il appelle volontiers les éléments du feu, de l'air et de la terre. Les planètes, une fois condensées, auraient été entraînées, et le sont encore, par cette giration de la « matière du ciel ».

Contre Descartes s'éleva Newton, qui s'immortalisa en formulant la loi de la gravitation universelle, rendant inutiles les tourbillons de Descartes. Mais la loi d'attraction de Newton implique un postulat difficile à comprendre, celui d'action instantanée à distance (cette difficulté a été levée en notre siècle par la théorie de la relativité générale).

Emmanuel Kant, le célèbre philosophe allemand, émit au XVIIIe siècle une nouvelle théorie cosmogonique, que l'on enseigne encore dans les classes sous le nom de « théorie de Kant-Laplace », car l'astronome français Laplace, sans la connaître, imagina la même, à peu de chose près. Une nébuleuse primitive, formée d'un essaim de météores d'après Kant, gazeuse d'après Laplace, animée d'un mouvement initial de rotation uniforme, s'est condensée, donnant lieu à un proto-Soleil entouré d'une atmosphère et qui a continué à tourner; de ce fait, l'atmosphère s'est contractée en anneaux, qui se sont détachés et sont devenus des planètes.

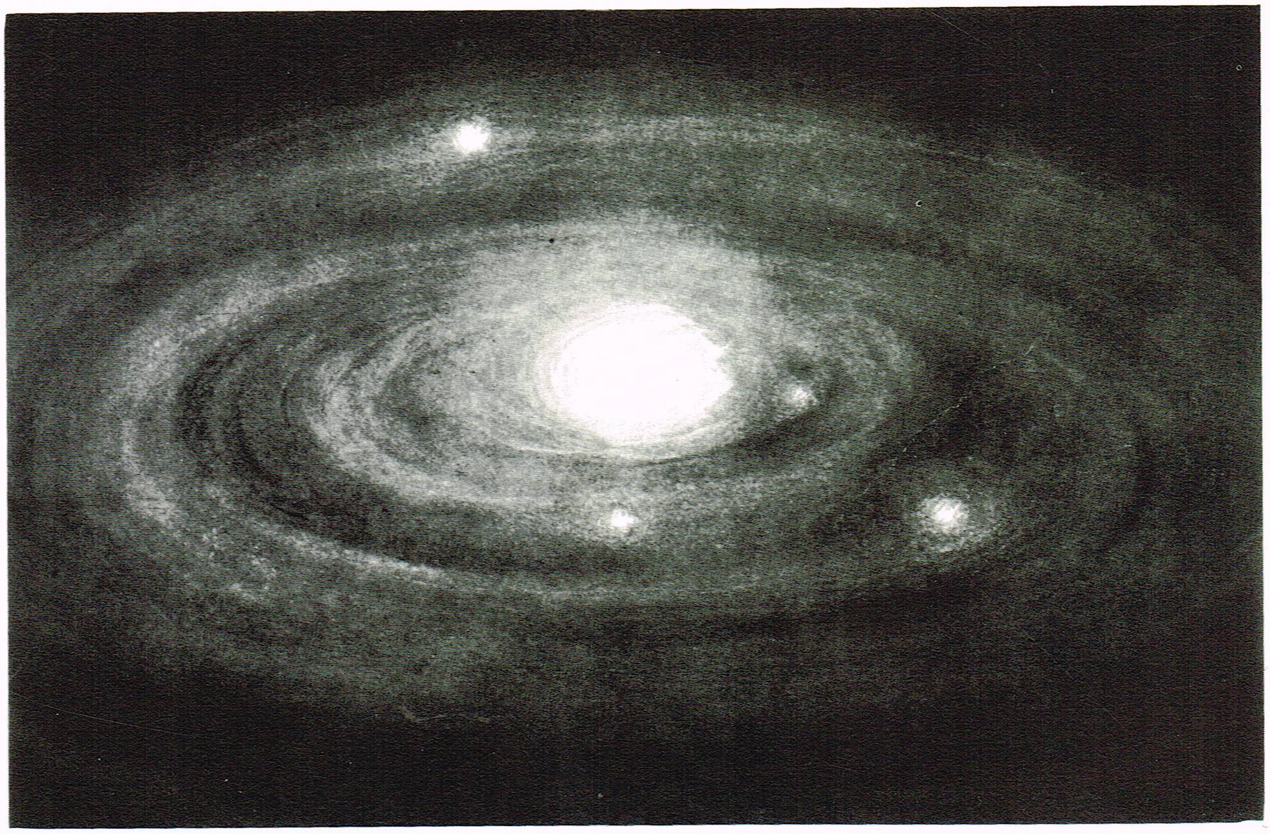
La théorie de Kant-Laplace a le mérite d'expliquer assez rationnellement pourquoi les orbites planétaires ont des rayons qui croissent, à mesure qu'on s'éloigne du Soleil, selon une progression géométrique (loi de Bode-Titius). Il faut cependant souligner que la planète Neptune, découverte en 1844, se trouve sensiblement plus proche du Soleil qu'il ne le faudrait et constitue aussi une exception à cette norme.

On a objecté à Kant et à Laplace qu'un Soleil formé comme ils l'indiquent devrait forcément tourner très vite; or, c'est le contraire qui a lieu : le Soleil tourne sur lui-même très lentement, en 27 jours. Les causes de ce ralentissement ont fait couler beaucoup d'encre, mais il y a plus grave : selon les travaux les plus récents sur la condensation, les anneaux de Kant-Laplace ne se condenseraient pas en corps célestes importants, mais en météores minuscules, peut-être même ne se condenseraient-ils pas du tout.

la Terre, fruit d'un cataclysme cosmique ?

Tournons-nous à présent vers un autre genre de théories cosmogoniques : les théories de la collision ou théories « catastrophiques ».

L'hypothèse cosmogonique de Laplace.



La première d'entre elles fut émise en 1746 par le naturaliste Buffon. Il imagina qu'une comète était passée assez près du proto-Soleil pour lui arracher des globes de feu qui devinrent les planètes. L'idée fut totalement abandonnée au profit de celles de Kant et Laplace, mais on la vit resurgir, très perfectionnée, au début de notre siècle. Parmi ces théories catastrophiques de notre temps, citons celles de Moulton et Chamberlin, Jeans, Jeffreys, etc.

Ces auteurs suggèrent qu'une étoile a dû passer près du Soleil et lui arracher un filament de matière, lequel s'est condensé en planètes. Malheureusement, cette thèse ne fournit aucune explication à la loi de Bode-Titius ni aux autres régularités du système solaire : orbites presque circulaires et toutes dans le même plan, etc.

Les anciennes théories catastrophiques, du type Buffon, ont une probabilité pratiquement nulle, étant donné l'immense distance qui sépare les étoiles les unes des autres. Les théories récentes supposent, au contraire, que ce choc a eu lieu, initialement, au centre de la Galaxie, quand les étoiles étaient beaucoup plus rapprochées; le système se serait déplacé après ce cataclysme.

Mais d'autres difficultés subsistent malgré l'élégance des mécanismes proposés par les auteurs : on est obligé de conclure qu'aucune théorie ne rend encore parfaitement compte de la formation du système solaire; mais l'importance de ces études cosmogoniques ne saurait être niée. Comme l'écrit E. Schatzmann : « Toute contribution aux théories cosmogoniques fait avancer notre intelligence de l'univers, et les théories cosmogoniques interviennent de façon dominante dans l'orientation de la recherche. »

notre croûte terrestre : trois milliards d'années

Quand on veut se limiter plus modestement à l'histoire de notre petit globe terrestre, on y trouve des certitudes plus substantielles. Que la Terre se soit formée par la condensation d'une masse gazeuse, par capture gravitationnelle de particules ou de toute autre manière, il reste qu'on peut fixer avec une grande probabilité la date de cette formation.

Pour déterminer l'âge de la Terre, les géologues se fondaient naguère sur l'épaisseur des terrains sédimentaires et les physiciens sur le refroidissement du globe. Ils aboutissaient à de fortes contradictions. Mais la découverte de la radioactivité a permis des mesures plus directes.

On connaît trois isotopes de l'uranium, qui se désintègrent en trois isotopes du plomb avec dégagement d'hélium. La période de désintégration est connue pour chacune de ces paires d'isotopes « uranium-plomb » portant des nombres atomiques différents. En déterminant l'abondance relative, actuelle, de chacun de ces isotopes du plomb dans une roche donnée, on arrive à en déduire son âge.

D'autres études se font sur les minerais de plomb et sur le potassium 40. De toutes manières, on arrive pour la Terre à un âge de l'ordre de trois milliards et demi d'années.

Que signifie, en réalité, ce nombre?

Il signifie que, depuis ce temps, la croûte terrestre existe approximativement sous la forme solide que nous lui connaissons; mais il ne laisse rien préjuger du temps qu'ont duré les phases précédentes de l'évolution terrestre, où notre globe existait déjà, mais à l'état liquide ou gazeux. Il est donc presque impossible, dans l'état actuel des connaissances, d'inférer l'âge du système solaire, ou même de tout l'univers à partir de celui de la Terre — en admettant que de telles questions aient un sens.

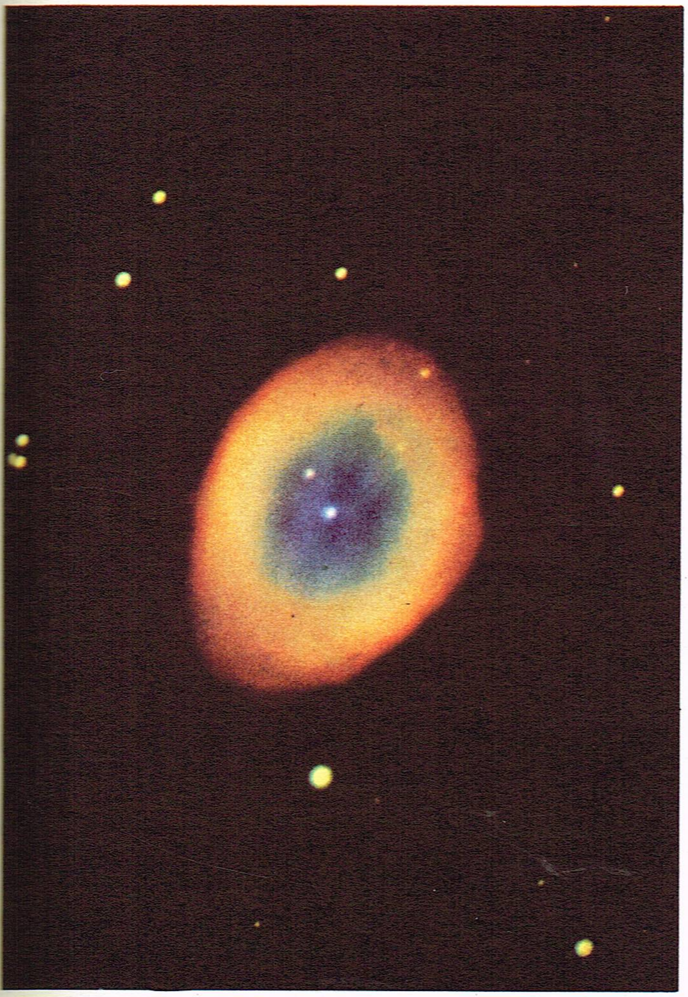
En réalité, des réponses ont été données, mais à partir d'études sur les étoiles et les nébuleuses, en quelque sorte par analogie. Les théories cosmogoniques feraient un grand progrès le jour où les astronomes pourraient assister à la formation d'une ou de plusieurs planètes dans le voisinage d'une étoile. En somme, et paradoxalement, c'est des objets les plus lointains et les plus difficilement accessibles que nous espérons recevoir des solutions aux problèmes qui nous intéressent le plus : ceux de nos origines et de notre destination.

En conclusion, précisons qu'on ne connaît pas encore exactement le mécanisme pouvant donner une explication de la genèse du système solaire (tout au plus sait-on qu'il existe), mais il ne semble pas qu'il y ait lieu de faire appel à des théories catastrophiques, dont le principe semble définitivement condamné comme étant en opposition absolue avec la mentalité scientifique, hostile à voir dans tout phénomène existant, et même existant à un très grand nombre d'exemplaires, le résultat du hasard. Du reste, théorie et observations sont d'accord pour considérer un système multiple ou planétaire comme normal et non accidentel (60 p. 100 des astres proches sont multiples, et ce pourcentage ne fait que croître avec les progrès de l'observation). On voit l'importance d'une telle affirmation : les « terres » sont innombrables et se comptent certainement par milliards de milliards dans notre Galaxie. Sont-elles habitables, et habitées?



Nébuleuse trifide (NGC 6514 ou Messier 20) du Sagittaire.

Nébuleuse du Crabe (NGC 1952 ou Messier 1) dans le Taureau.



Nébuleuse annulaire de la Lyre (NGC 6720 ou Messier 57).



(Phot. California Inst. of Technology.)

des mondes habitables, et habités ?

Existe-t-il, dans l'univers, des globes analogues à notre Terre et capables de servir d'habitat à des végétaux, à des animaux, voire à des êtres pensants comparables à l'homme? Cette question s'est posée à nous de tout temps, mais, jusqu'à notre époque, nul n'y pouvait répondre que par des suppositions gratuites, le plus souvent fantaisistes. Dans l'Antiquité, Lucrèce la résolvait déjà par l'affirmative : « Conviens que, dans le reste du vide, les éléments de la matière ont enfanté sans nombre des êtres animés, des mers, des cieux, des terres, des mondes semblables à celui qui se balance sous nos pas dans les flots aériens! »

On ne saurait dire mieux, et, somme toute, nous ne sommes pas tellement plus avancés sur ce point que le poète antique.

La science, à notre époque, ne peut pas décider encore si les planètes du système solaire sont, ou ne sont pas, habitées, mais les progrès impressionnants de la cosmonautique permettent d'espérer pour bientôt une réponse à cette question. Quant aux étoiles lointaines, on est parvenu à établir, comme nous le verrons plus loin, qu'elles ont souvent des planètes, et que rien ne s'oppose à ce que la vie se soit développée sur un grand nombre d'entre ces « corps obscurs ».

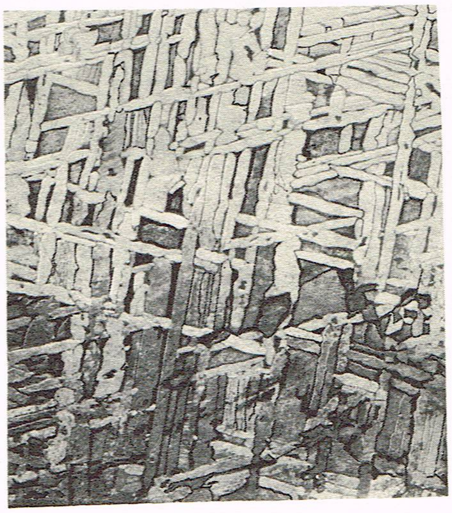
Nous examinerons successivement les deux parties de ce problème : système solaire et étoiles; mais, auparavant, il est nécessaire de préciser quelques notions générales sur la vie.

Lorsqu'on parle, en science, de la vie dans l'univers extra-terrestre, on entend une vie analogue à celle que nous connaissons sur la Terre. Il n'est pas impossible d'envisager des formes de vie très différentes, et les écrivains de science-fiction n'y manquent pas. L'hypothèse la plus sérieuse consisterait à supposer, par exemple, des organismes où le silicium remplacerait le carbone (nous y reviendrons plus loin). On pourrait également penser à des êtres vivants, et même pensants, absolument imperceptibles à l'homme, bien que matériels. Voltaire y avait déjà songé (Micromégas). Mais on peut aussi rêver à des êtres à deux dimensions, etc. Bornons-nous à indiquer quelles sont les conditions matérielles nécessaires à une vie pareille à celle que nous connaissons sur notre Terre.

Essayons donc de raconter par quels processus successifs la vie est apparue et s'est développée ici-bas. C'est une vaste énigme, que la biologie, la chimie, la paléontologie et la géologie s'efforcent de résoudre depuis environ cent cinquante ans...

Une théorie qui eut une grande vogue, et qui compte encore des partisans, est celle de la « panspermie cosmique ». On supposa d'abord que la Terre avait été ensemencée de germes vivants par les météorites, ce qui se révéla, par la suite, impossible, vu la haute température de ceux-ci. Mais le savant suédois Svante Arrhenius, il y a plus d'un demi-siècle, développa une théorie suivant laquelle la Terre aurait été ensemencée à partir du système solaire, grâce à la « pression de radiation » du Soleil.

Toutes ces hypothèses visent à écarter la nécessité d'une génération spontanée initiale — idée qui répugne particulièrement aux savants depuis les belles expériences de Pasteur. Malheureusement, cette génération se trouve simplement reculée, dans cette théorie, aux planètes du système solaire. Il n'est pas possible de concevoir un ensemencement de ces dernières par les étoiles, bien trop lointaines! En outre, le transport des germes, de planète à planète, semble présenter des difficultés insurmontables : froid extrême qui règne dans l'espace interplanétaire et, plus encore, action nocive des radiations ultraviolettes et des rayons cosmiques, stérilisant ces germes.





La météorite qui tomba, en mai 1864, à Orgueil, près de Montauban — dont on voit ici un fragment et sa coupe —, a été récemment réétudiée, aux États-Unis, par le professeur Nagy, qui y a trouvé des vestiges de cellules organiques et même tout un organisme microscopique d'une forme inconnue sur la Terre. Certains contestent cette observation, et, d'ailleurs, ajoutent-ils, on ignore la nature et la provenance des météorites...

les origines de la vie

On se résigne donc à admettre que la vie terrestre a pris naissance sur la Terre même, lorsque celle-ci a été suffisamment refroidie. On postule que toutes les molécules actuellement contenues dans les cellules vivantes existaient sur la Terre dès le début. N'existent-elles pas aussi sur les autres planètes et même dans tout l'ensemble de l'univers connu ? Aux premiers âges, l'atome de carbone joua, de toute évidence, un rôle prépondérant dans la genèse des substances organiques. Car il avait seul le pouvoir exceptionnel de former des macromolécules en s'unissant à l'hydrogène, à l'oxygène et à l'azote. (Ces grosses molécules organiques comportent jusqu'à des milliers d'atomes.) Il semble qu'aux premiers âges de la Terre le carbone n'existait que sous forme de gaz carbonique; l'atmosphère ne contenant pas d'oxygène, donc pas d'ozone, sol et mer étaient exposés aux rayons cosmiques et aux radiations X et ultraviolettes de très haute énergie (théorie de Dauvilliers). De telles conditions semblent favorables à une création d'aldéhyde formique selon une réaction qui fut réalisée en laboratoire, dans des conditions apparemment analogues à celles que présentaient les couches superficielles de la planète Terre :

$CO_2 + H_2O \rightarrow O_2 + HCHO$

qui explique la formation de l'oxygène (puis de l'ozone) de l'air, puis

$6\ HCHO \rightarrow C_6H_{12}O_6$ (glucose).

On ne sait pas encore exactement comment les molécules et les chaînes organiques naquirent, ni comment la vie — d'abord sous une forme élémentaire, puis sous des formes de plus en plus complexes, jusqu'à celle de l'homme — prit définitivement droit de cité. Au laboratoire, il est facile de passer de l'aldéhyde formique aux sucres, puis, par combinaison avec l'azote atmosphérique, aux protéines, le tout dans des vases où règnent des conditions et une atmosphère comparables à celles de la Terre primitive.

Ainsi se seraient constitués des vastes bancs de gelée protéique dans les lagunes saumâtres qui recouvraient alors la plus grande partie du globe terrestre. (La vie ainsi créée est très sensible aux radiations ultraviolettes, mortelles pour elle. Heureusement, la biosphère dégage de l'oxygène, qui, sous l'action des radiations ultraviolettes, crée la couche infime — de quelques centimètres d'épaisseur — d'ozone qui nous protège encore.)

Même si l'on tient compte des milliards d'années qui séparèrent vraisemblablement la genèse de la première macromolécule organique de la genèse de la première algue, puis du premier animal, voire du premier homme, il est plus difficile d'expliquer la différenciation progressive des êtres organisés, et notamment leur nutrition et leur reproduction; certains paléontologues veulent même faire intervenir un principe directeur, qui aurait orienté les « hasards heureux » et donné le sens de l'orthogenèse aboutissant à l'homme. Mais, ici, nous entrons dans un domaine philosophique qui sort du cadre de cette étude.

les conditions de la survie

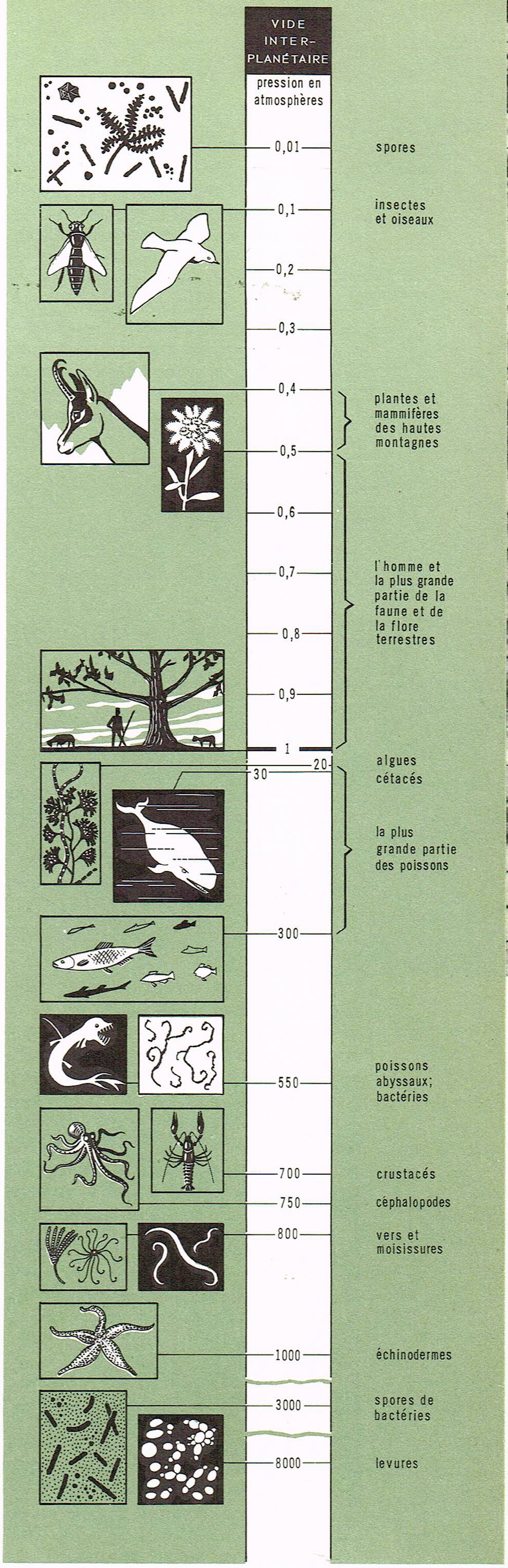
Le phénomène « vie » semble très vivace, voire indestructible, une fois créé : les conditions en Sibérie, dans le fond des fosses océaniques et au cœur de la savane africaine semblent bien diverses; de même, la survie de germes de blé déposés il y a des millénaires dans les tombes de la vallée des Rois, ou des semences enfouies dans la glace polaire, a de quoi laisser rêveur sur les possibilités de la molécule vivante. Il semble cependant que les conditions indispensables pour que de semblables processus apparaissent sur d'autres mondes que le nôtre sont principalement les trois suivantes.

température convenable

La vie ne peut subsister que dans des limites de température relativement restreintes. On admet généralement que le protoplasme est tué à + 100 °C, bien qu'il y ait quelques exceptions pour des spores ou autres organes de conservation des organismes les plus primitifs. On pense que toute activité vitale cesse bien avant le zéro absolu. Une telle gamme de températures, si étroite par rapport à celle des états extra-terrestres de la matière, semble cependant normale sur les planètes.

présence d'eau

L'eau est absolument indispensable à la vie telle que nous la connaissons. Toute cellule en contient pour 60 à 90 p. 100 de son volume. C'est pourquoi l'on a qualifié parfois les organismes d' « aquariums ambulants » et on incline à leur attribuer une origine marine, leur



transfert « au sec » ne s'étant accompli que relativement tard. Cependant, les lichens et leur faune minuscule supportent aisément de longues sécheresses, grâce à leur pouvoir de dessiccation suivie de reviviscence. Qui ne connaît, parmi les plantes supérieures, l'exemple de la « rose de Jéricho »? Il faut dire cependant qu'aucune nutrition n'est possible sans l'eau, qui dissout les substances alimentaires; que l'eau joue un rôle important de régulation thermique et évite les grands écarts de température, mortels aux organismes. Aussi le problème de l'existence de l'eau sur les planètes, et notamment sur Mars, est-il si important, quoique difficile à résoudre. L'atmosphère terrestre en est riche, et seul un observateur qui se trouverait au-delà peut espérer déceler les quantités d'eau, de toute manière infimes, et sa pression sur la planète voisine.

atmosphère respirable

Cette condition va de soi, bien que de nombreux organismes soient capables de vivre en l'absence complète d'oxygène (anaérobiose), Mais l'atmosphère n'en fournit pas moins l'oxygène et le gaz carbonique, éléments essentiels du cycle vital. En outre, c'est la pression atmosphérique qui maintient l'eau à l'état liquide : sans atmosphère, aucune mer, aucun cours d'eau. Elle protège également le sol contre l'impact des météorites, qui, on le croit, ont ravagé par exemple la surface de la Lune.

Or, aucun corps céleste de petite taille ne peut retenir une atmosphère autour de lui — cela pour des raisons physiques qu'il serait un peu long d'exposer ici (et, notamment, par suite du manque d'attraction gravitationnelle, alors que l'agitation désordonnée des molécules à température non nulle favorise leur évasion). A l'instar de notre Lune, les satellites des planètes en sont certainement dépourvus, ainsi que la totalité des astéroïdes.

Passons maintenant en revue les planètes de notre système. Voyons si la vie est possible à leur surface, et précisons quelle forme de vie on pourrait éventuellement s'attendre à y trouver.

Mercure

C'est la planète la plus proche du Soleil. Elle est à peine plus grosse que la Lune. Si l'on se rappelle ce qui vient d'être dit, Mercure ne peut avoir ni atmosphère ni eau — ce que l'observation confirme d'ailleurs, bien qu'on ait décelé autour d'elle une mince pellicule gazeuse, probablement d'argon. Mercure tourne toujours la même face du côté du Soleil, d'où il résulte que cette face est torride, tandis que l'autre reste éternellement obscure et glacée. Ce sont là toutes conditions qui excluent la vie telle que nous la connaissons.

Vénus

Elle vient en deuxième position à partir du Soleil. Elle se trouve entre Mercure et la Terre, et la distance qui la sépare de nous ne paraît plus, de nos jours, infranchissable. Mais quels habitants trouverons-nous sur Vénus? Cette planète, beaucoup plus grosse que Mercure, a presque le même diamètre que la Terre, et à peu près la même densité. Comme Mercure, elle semble toujours tourner la même face vers le Soleil. Il semble, à en juger d'après sa densité, qu'elle soit recouverte d'une croûte minérale solide (de roches analogues au granite). Mais nous connaissons mal Vénus, car sa surface est perpétuellement recouverte d'une « mer de nuages » qui la dérobe à nos investigations. On pense que la température moyenne de son hémisphère éclairé est élevée, de l'ordre de 400 °C, et que, dans son hémisphère obscur, cette température ne descend pas au-dessous de — 30 °C. Ce sont, on le voit, des conditions peu favorables à la vie. De plus, l'atmosphère de Vénus ne paraît contenir ni oxygène ni vapeur d'eau, et ses nuages seraient faits de microcristaux d'aldéhyde formique. Cependant, cette atmosphère est 160 fois plus riche que la nôtre en gaz carbonique. On reconnaît les conditions que nous avons décrites plus haut comme étant celles de la Terre aux premiers âges, la vapeur d'eau en moins. Tout compte fait, **Vénus est probablement le type même d'une planète où la vie en est encore à ses débuts.** Il faut se rappeler qu'elle a été détachée du Soleil après la Terre — elle est « jeune ». Mercure, encore plus récent, est pour ainsi dire à l'état de nouveau-né. Si, de ce point de vue, on considère la Terre comme une planète adulte, Mars, en revanche, détaché avant elle, est déjà « vieux ».

On risque donc de trouver sur Vénus une vie végétale abondante, mais peut-être encore rudimentaire, et pas ou peu d'animaux. Espérons que l'exploration astronautique nous fixera bientôt sur ce point, comme sur bien d'autres.

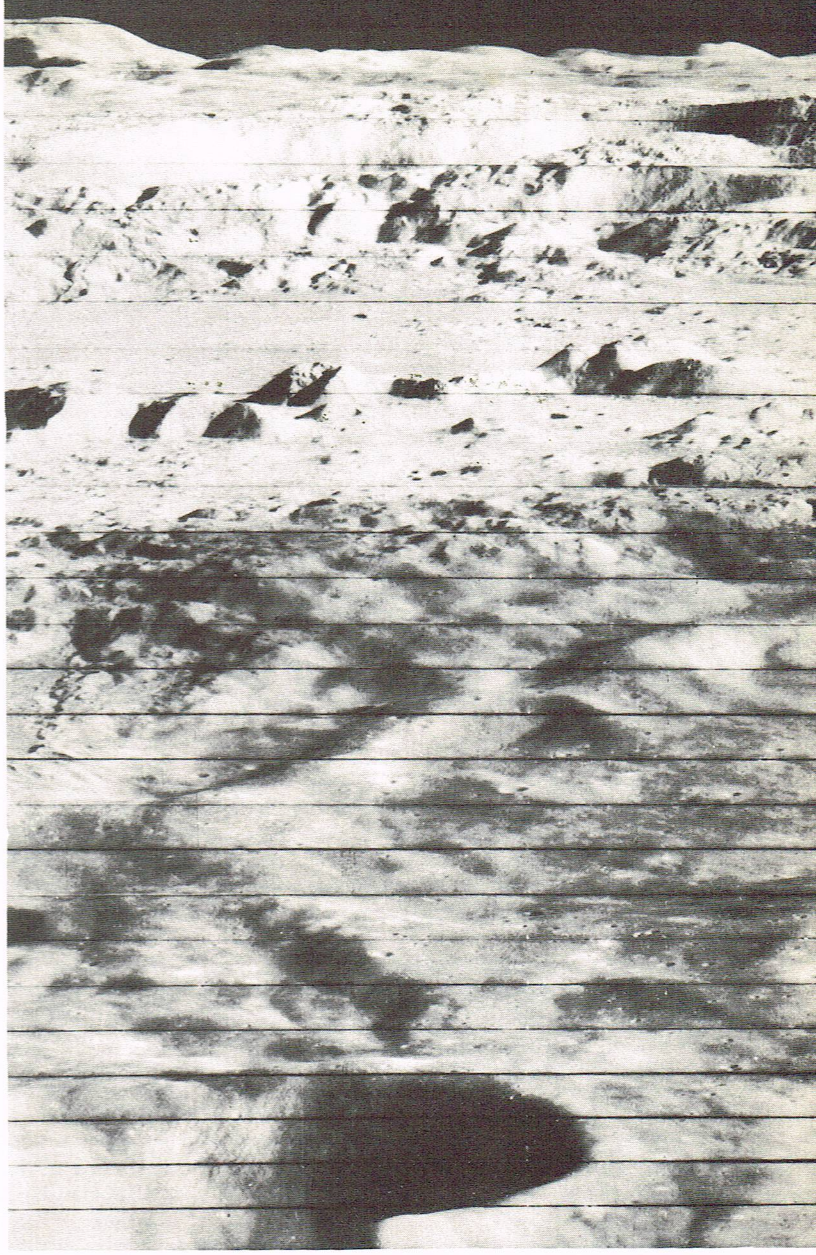
la Lune

Notre satellite nous est beaucoup mieux connu. Il présente peu de différences entre sa face visible, objet d'un nombre considérable d'études, et sa face invisible, que la cosmonautique a déjà pu explorer grâce à des instruments portés par un satellite artificiel. Elle nous présente toujours la même face, par suite de l'interaction mécanique entre nous et elle, qui a égalisé rapidement les vitesses de rotation. Son volume ne dépasse guère 2 p. 100 de celui de la Terre. Comme prévu, il n'y a pas d'atmosphère, ou tout au plus une atmosphère raréfiée à l'extrême (les gaz s'en sont depuis longtemps échappés, l'attraction gravitationnelle étant trop faible pour les retenir). L'eau manque totalement, depuis longtemps évaporée elle aussi. Les prétendues « mers lunaires » ne sont que des étendues de sable, de gravier ou de cendre. C'est un « monde mort » dans toute l'acception du mot. Les écarts de température sont énormes entre le jour et la nuit lunaires (qui durent respectivement 14 jours). Aucune vie ne peut prospérer dans des conditions aussi défavorables.

Mars

Il vient, à partir du Soleil, en quatrième position, après la Terre (la Lune ne compte pas dans la série, elle fait partie du système « Terre »). Mars nous est beaucoup mieux connu que Vénus, car sa surface est nettement visible. Il est sensiblement plus petit que la Terre et il possède deux satellites. Sa densité, proche de la nôtre, permet de conclure à une croûte superficielle solide. Mars connaît l'alternative des jours et des nuits, car il tourne sur lui-même en 24 h 37 mn. Il a également des saisons, vu l'inclinaison de son axe de rotation sur le plan de son orbite. On le voit, ce monde ressemble singulièrement au nôtre. **Si la vie existe ailleurs dans le système solaire, c'est sur Mars, certainement, qu'on peut s'attendre à la rencontrer. Il faut espérer que les progrès foudroyants de l'astronautique nous permettront bientôt (mais à quel prix?) d'y aller voir, ce qui est plus sûr que de spéculer.** Le trait qui frappe tout de suite celui qui regarde Mars dans un télescope, c'est la présence de « calottes polaires » d'une blancheur de neige. Ces calottes, comme on peut le prévoir, disparaissent au début de l'été martien. Avec les saisons, les « mers » de Mars changent de coloration. On ignore exactement la nature de ces « mers », et l'on présume qu'elles ne

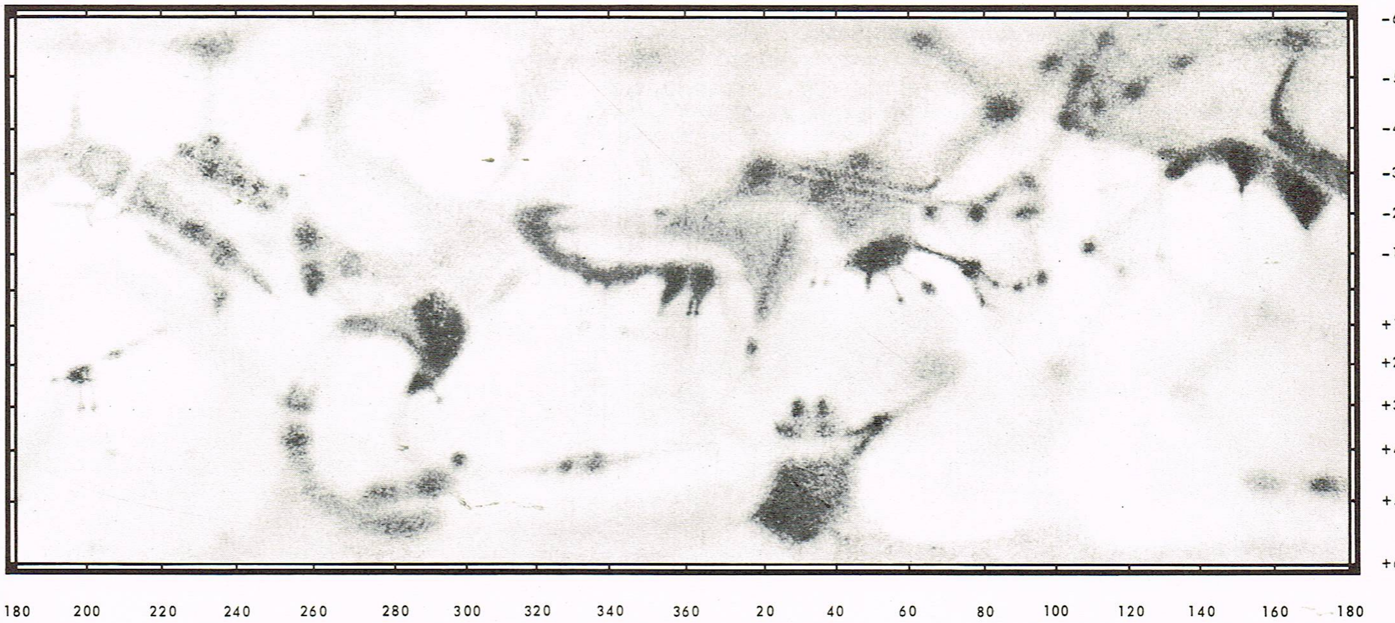




Ci-dessus, les photographies prises par la sonde « Mariner IV » surprirent en révélant que le sol martien est criblé de cratères comme l'est celui de la Lune. *(Doc. Jet propulsion Lab. Calif. Inst. of Tech.)*

Ci-contre, la caméra de la sonde « Lunar Orbiter II » a obtenu cette vue plongeante du cirque Copernic sur la Lune. Au fond, la chaîne des Carpathes. *(Doc. Observatoire du Mont-Wilson. NASA.)*

Ci-dessous, ce planisphère de la planète Mars synthétise les observations de Slipher, Lyot, Camichel, Gentili et G. de Vaucouleurs.



Doc. Observatoire du Mont-Wilson.

contiennent pas d'eau; leur surface serait recouverte de végétaux inférieurs, qui changent de couleur entre l'hiver et l'été. Plutôt que de lichens, il pourrait s'agir d'organismes microscopiques analogues à nos bactéries chromogènes. Les zones claires, ou « déserts », de Mars seraient nues et constituées d'un oxyde de fer.

Il nous reste à parler des célèbres « canaux » de Mars. Ces stries rectilignes, le plus souvent géminées, ont pu faire croire à la présence actuelle ou passée d'une haute civilisation, responsable de ces ouvrages d'art. Mais, après des discussions passionnées et qui durent encore, force a été d'abandonner cette hypothèse et de reconnaître que les canaux de Mars n'existent pas : ce sont des illusions d'optique dues à des instruments insuffisants.

L'atmosphère martienne, très pauvre en oxygène et vapeur d'eau, paraît impropre à la respiration des animaux et de tout être supérieur. Mais qui sait ?...

les astéroïdes

Ces très petites planètes, qui gravitent en essaim entre Mars et Jupiter, ne sont que des blocs de rochers dont le plus gros n'a même pas 800 km de diamètre; ni atmosphère ni eau dans ces conditions. La température doit y être très basse et impropre à la vie.

Jupiter

C'est une sphère énorme et de faible densité, dont le volume fait plus de mille fois celui de la Terre. Cette planète, déjà très éloignée du Soleil, est extrêmement froide. On la croit entourée d'une épaisse atmosphère de méthane, ammoniac, hydrogène, azote et hélium. Au-dessous, ce sont probablement des couches liquides, puis un noyau glacé. Aucune vie ne peut exister dans de pareilles conditions.

Saturne, Uranus, Neptune, Pluton

Le froid augmente à mesure qu'on s'éloigne du Soleil. Il semble bien qu'aucune de ces planètes ne puisse donner un quelconque refuge à des êtres vivants.

Il semble bien qu'il ne puisse guère exister dans le système solaire qu'une vie végétale inférieure, sur Vénus et sur Mars. En est-il de même dans la Galaxie qui nous entoure et dans les autres galaxies de l'univers ?...

combien de planètes habitées dans toute la Galaxie ?

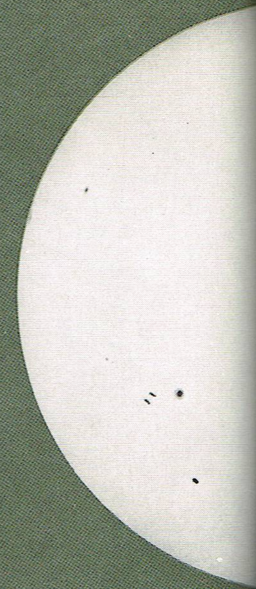
Ici, le nombre de chances pour qu'il existe une ou plusieurs planètes biologiquement semblables à la Terre, si réduit dans le système solaire, se trouve multiplié par des millions, voire par des milliards... Mais précisons un peu : rien que dans notre Galaxie il peut exister plusieurs centaines de millions de systèmes planétaires, même si toutes les étoiles n'ont pas des planètes, certaines n'ayant pas encore atteint le stade de ce morcellement, d'autres l'ayant dépassé depuis longtemps et résorbé les planètes qu'elles ont pu avoir dans un passé lointain.

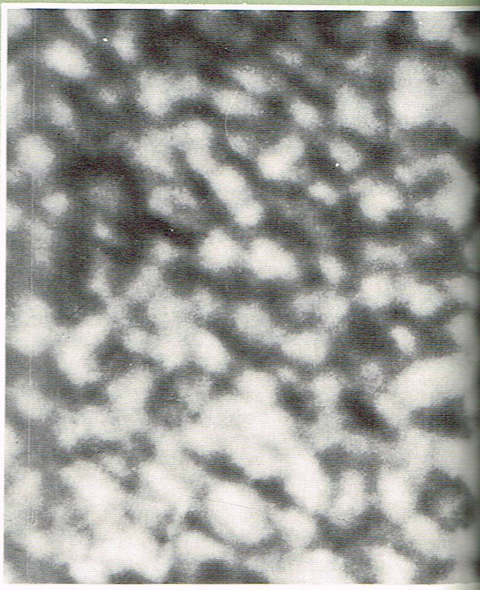
Tous les systèmes planétaires ne sont pas analogues au nôtre. Certains ne comportent apparemment qu'une planète unique, de grande taille. D'autres en ont plusieurs, mais les unes sont trop rapprochées du foyer central, donc torrides ou grillées, les autres en sont trop éloignées, donc glacées. Il y a aussi des systèmes où les orbites sont si excentriques que les planètes passent par d'extraordinaires écarts de température. Dans les systèmes d'étoiles doubles (et ils sont très nombreux), les planètes sont également amenées à passer bien trop près des foyers. **Ces restrictions étant faites, on peut estimer à quelques millions le nombre des systèmes planétaires que notre Galaxie renferme.** C'est un nombre fabuleux, qui a pu être précisé grâce à des travaux récents. Disons brièvement qu'on arrive à détecter la présence de planètes dans le voisinage d'une étoile non par l'observation directe (l'éclat de l'étoile « noie » absolument les corps obscurs qui l'avoisinent), mais par l'étude des perturbations que provoquent ces compagnons dans les mouvements de leurs « soleils ».

Or, notre Galaxie n'est qu'une infime partie de l'immense univers, où l'on a dénombré déjà un milliard d'autres galaxies. **C'est dire que les chances de la vie semblent positivement énormes, et il est fort difficile de croire qu'en vertu d'un mystérieux arrêt elle a été strictement réservée à notre minuscule demeure.**

Bien entendu, ce n'est pas encore demain que nous le saurons (l'étoile la plus proche, α Proxima du Centaure, est à quatre ans de voyage pour une fusée naviguant à la vitesse de la lumière), et force nous est de spéculer qu'un phénomène réalisé sur la Terre est plausible ailleurs.

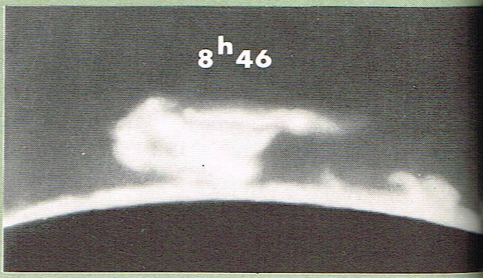
De gauche à droite, **le Soleil photographié en lumière blanche et, au spectrohéliographe, en lumière monochromatique : raies α de l'hydrogène et K du calcium.**





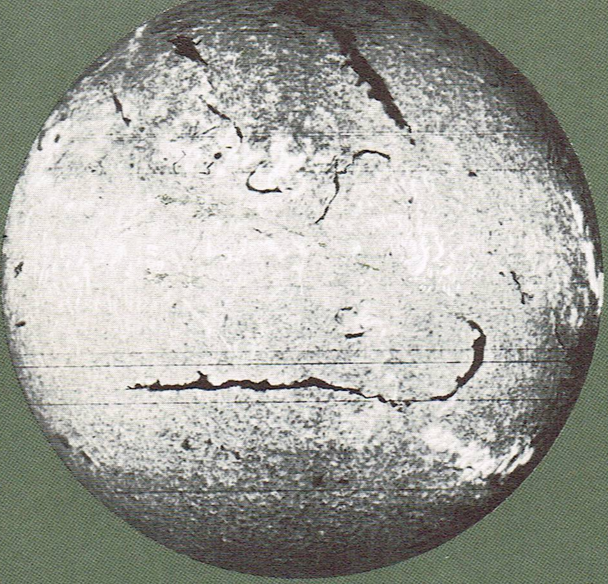
Les « grains de riz » photographiés à 28 000 m, hors des couches denses de l'atmosphère, par un ballon porteur d'un télescope.

Ci-dessous et page suivante, **évolution d'une protubérance éruptive.**













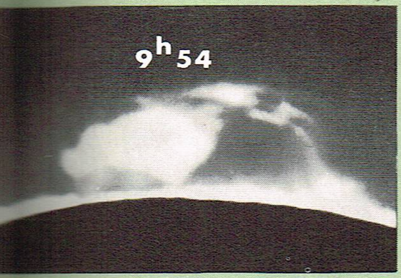
le Soleil

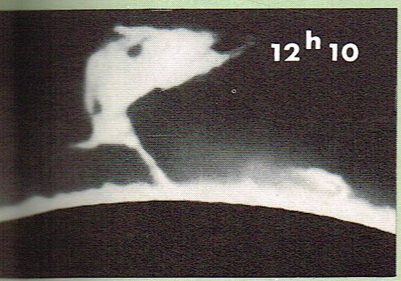
Le Soleil, étoile la plus proche, est aussi la mieux connue. Il nous apparaît sous un diamètre apparent assez grand pour que l'étude des détails de sa surface soit possible et même relativement facile. Mais c'est la luminosité considérable du disque qui gêne souvent cette étude. Il est d'usage de distinguer les propriétés du « Soleil calme » et celles qu'il présente pendant ses périodes d'activité électromagnétique, responsables de ces fameux « orages » magnétiques dont les effets se font sentir jusque sur la Terre.

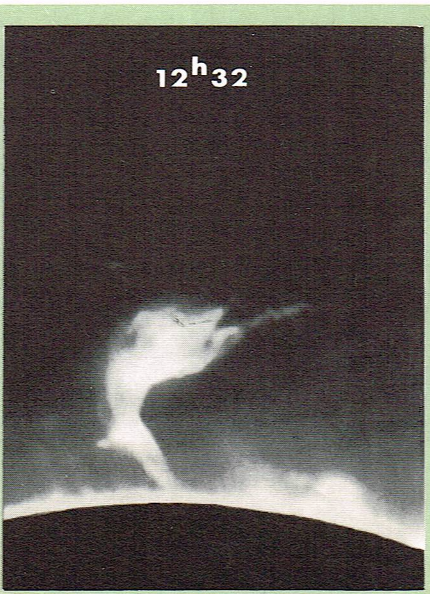
En l'absence de cette activité, le disque solaire présente un certain nombre de particularités constantes. La photographie de sa surface et l'observation visuelle mettent en évidence la présence d'irrégularités appelées « grains de riz »; c'est la « granulation », qui évolue constamment : les grains disparaissent au bout de quelques minutes et laissent la place à d'autres. On pense que la granulation serait due à des mouvements de convection se produisant à quelque profondeur dans l'étoile. Quant à la surface elle-même, on ne peut l'observer que sur les bords et à condition que le disque soit masqué, car sa trop grande luminosité rend toute observation impossible. Aussi cette étude n'a-t-elle été longtemps possible que pendant les rares éclipses totales de Soleil : en effet, les diamètres apparents de la Lune et du Soleil étant sensiblement égaux, le disque solaire peut être caché par la Lune.

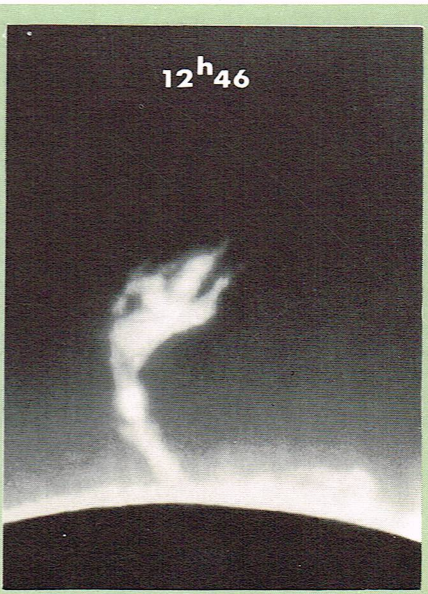
On observe alors, au voisinage immédiat du Soleil, une première enveloppe, relativement mince, de couleur rose, la chromosphère, qui est hérissée de flammèches roses, d'aspect variable : les protubérances. Au-delà de la chromosphère apparaît une seconde enveloppe, comparable à une auréole blanche, qui environne le soleil éclipsé : la couronne. La couleur rosée de la chromosphère est due à une forte émission d'une raie de l'hydrogène : $H\alpha$. Cette particularité a été mise à profit pour étudier cette première enveloppe, même en dehors des éclipses, selon le principe du spectrohéliographe, qui a beaucoup ajouté à nos connaissances dans les premières décennies de ce siècle (L. d'Azambuja).

L'étude de la couronne est plus difficile, car elle n'émet pas de raie qui permette, comme $H\alpha$ dans le cas de la chromosphère, de l'isoler du reste du disque par des moyens











Deux beaux phénomènes dus à l'activité solaire : *ci-contre*, couronne solaire lors d'une éclipse totale; *à droite*, aurore australe.

spectrographiques. Elle n'est bien visible que pendant les éclipses totales; mais on sait qu'une éclipse de Soleil n'apparaît jamais comme totale en tous les points de la Terre en même temps; aussi, chaque fois qu'une éclipse est prévue, c'est un grand branle-bas des astronomes solaires, qui, avec armes et bagages, se transportent aux endroits les plus insolites.

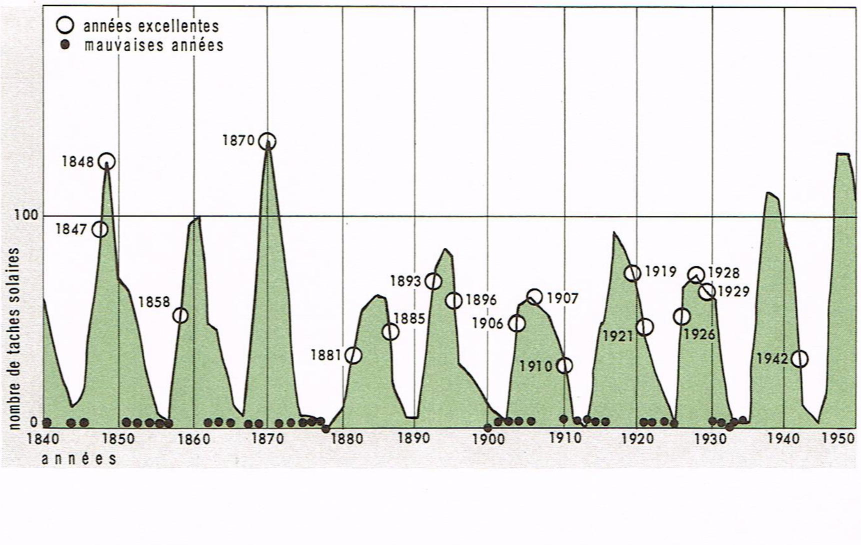
L'esprit d'invention d'un Français, Bernard Lyot, a permis de pallier en partie ces inconvénients; on peut en effet procéder à l'observation continue de la couronne solaire grâce à un appareil de son invention : le coronographe. Un écran intercepte les rayons lumineux en provenance du disque, et de grandes précautions sont prises pour que les rayons diffusés par l'écran soient absorbés avant de parvenir dans le système optique qui donne l'image de la couronne. Le rayonnement total de cette dernière n'est que le millionième de celui du Soleil; on peut juger des difficultés à surmonter.

L'étude du spectre de la couronne met en évidence des raies très intenses, dues à des atomes à un très grand degré d'ionisation : fer ionisé neuf et dix fois (ayant perdu neuf ou dix électrons); calcium ionisé onze et douze fois, et ainsi de suite. On s'est longtemps perdu en conjectures sur ces « raies coronales » qui ne semblaient correspondre à aucun élément commun. L'étude théorique de ces émissions indique paradoxalement une température extrêmement élevée, sans aucune comparaison avec la température superficielle, qui est de l'ordre de 6 000 °C.

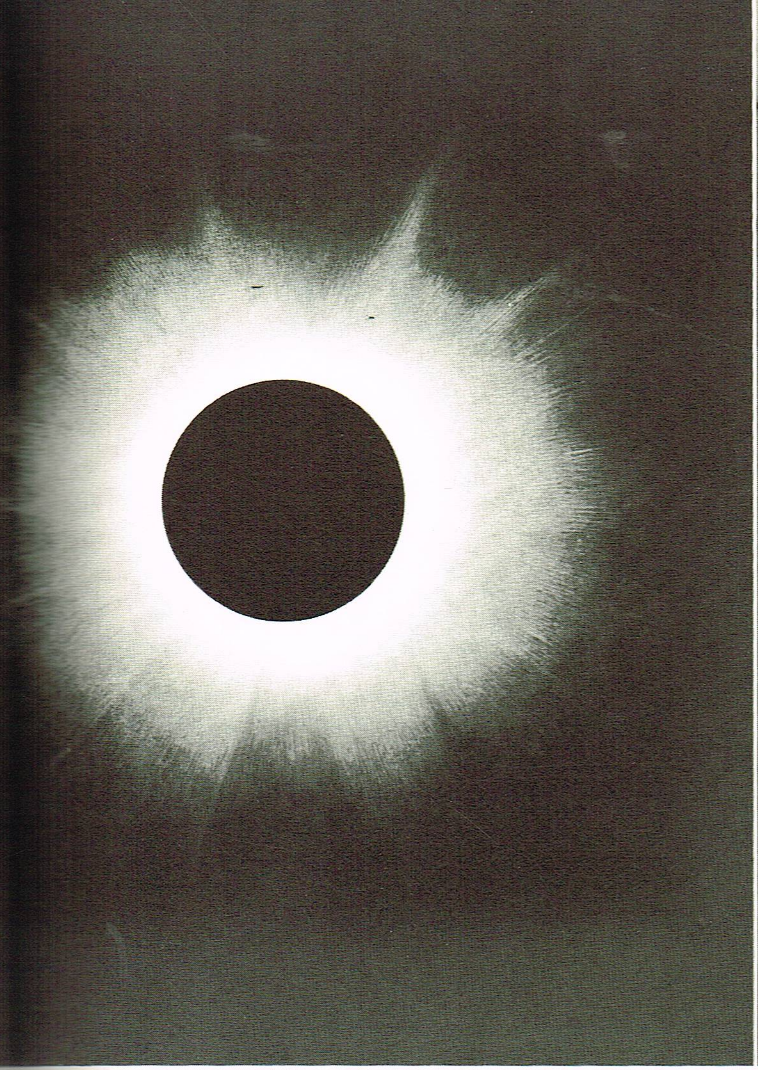
L'examen du degré d'ionisation des métaux et de l'émission des ondes radio révèle en effet une « température cinétique » des électrons de l'ordre du million de degrés. On a cherché à expliquer ces hautes températures en faisant intervenir des « ondes de compression » et même des « ondes de choc » qui monteraient de la chromosphère.

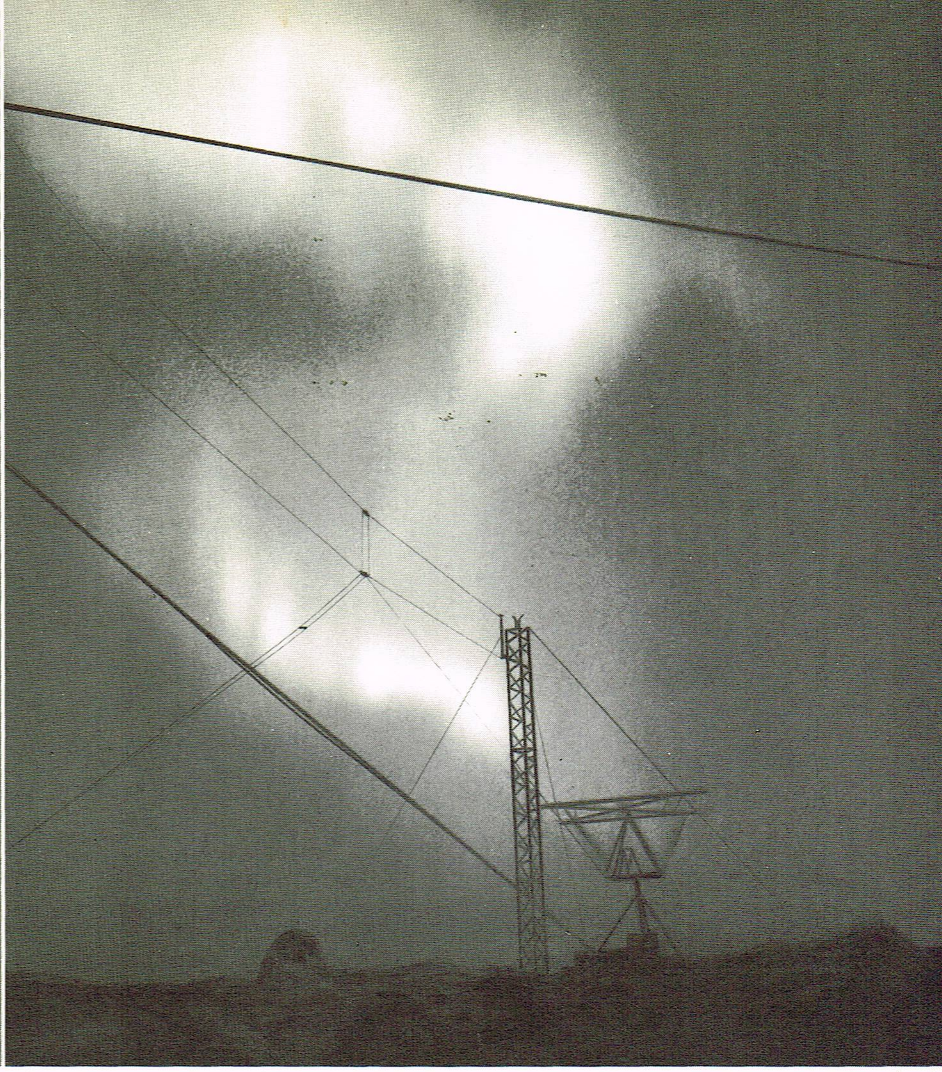
Un mot maintenant de cette activité solaire à laquelle la superstition populaire attache une puissance quelque peu maléfique : tremblements de terre, périodes de canicule insolites, révolutions et même des guerres lui sont attribués avec plus ou moins de fondement.
Pour l'astrophysicien, l'activité solaire se manifeste tout d'abord par l'apparition de taches sombres facilement observables avec une petite lunette munie d'un verre noirci. Effectivement, le nombre des taches présente une périodicité de onze ans, la période de grande activité solaire étant aussi celle où les taches sont le plus nombreuses : c'est ce qu'on appelle le « cycle solaire ».

Cette activité se manifeste aussi par un nombre plus élevé de protubérances, qui se présentent fréquemment sous la forme d'arches de faible épaisseur. On observe aussi souvent de véritables éruptions chromosphériques : la région intéressée devient très brillante, un flot de gaz à très haute température est rejeté par la surface solaire. Certains phénomènes terrestres (orages magnétiques, aurores boréales) sont liés à ces éruptions. Ils peuvent s'expliquer par une ionisation de la haute atmosphère terrestre sous l'effet des radiations ultraviolettes qui accompagnent les éruptions; l'arrivée des rayons cosmiques est modifiée, de même que la propagation des ondes électromagnétiques.



L'activité solaire d'après le nombre des taches solaires (d'après S.B. Nicholson). Nous avons indiqué la qualité des vins de Bourgogne pour la période 1838-1950 (d'après G. Bidault de l'Isle).





Pendant longtemps la première des sciences, l'astronomie s'était un peu endormie sur ses lauriers faute d'instruments suffisants d'observation.

Plusieurs « révolutions » devaient la bouleverser. Tout d'abord la « révolution astrophysique » avec la spectrographie, qui permet à l'homme de pénétrer au cœur même des étoiles, d'en analyser composition et structure. Nous avons vu comment un principe essentiel s'est dégagé au début de ce siècle : où que nous portions nos regards dans le monde, et même lorsque nous découvrons ici-bas de nouvelles substances, les mêmes substances et les mêmes forces existent et agissent dans tout l'univers!

Bien plus récemment l'homme émerveillé constatait que les astres sont d'immenses laboratoires où se déroulent depuis des millénaires des processus inconcevables ici-bas, ou peu s'en faut : physique des plasmas, températures de plusieurs millions de degrés, gaz à très haut degré d'ionisation, états « super-denses », réactions nucléaires.

Puis, avec la « révolution cosmologique » s'ouvrit un prodigieux foyer d'intérêt et de recherche, né du besoin qu'a l'homme de situer son modeste logis dans un ensemble de plus en plus vaste et ordonné.

Enfin, plus jeune, mais combien populaire, la « révolution astronautique » autorise les plus grands espoirs pour demain. Grâce aux projectiles extra-terrestres, l'astronomie se débarrassera de la « gangue » que constitue l'atmosphère terrestre. Quelle moisson scientifique ramènera ou transmettra à la Terre le premier astronome (ou le premier instrument) qui pourra s'échapper dans le vide interplanétaire! Les progrès techniques suscités par les besoins des astronautes servent aussi largement à l'astronomie fondamentale ; pensons aux gigantesques radiotélescopes, dont il sera parlé au prochain chapitre, et qui servent aussi bien à suivre « à la trace » les satellites qu'à détecter des galaxies inconnues aux confins de l'univers. Mais pensons surtout à l'immense regain de faveur et d'intérêt de l'astronomie chez les jeunes qui, dès l'âge de « Tintin », ne rêvent que d'expéditions intersidérales. Il n'est pas douteux que, pour eux, l'astronomie sera, avec la physique théorique, un des grands pôles d'attraction de la science de demain.

LA RADIO

regards nouveaux sur l'univers



Radiotélescope de 25 m de Dwingeloo (Pays-Bas). L'appareillage radio-astronomique de l'observatoire de Nançay a été présenté pages 21 à 28.

L'instrument principal de l'observatoire radio-astronomique de Nançay est maintenant l'un des plus grands et l'un des plus précis des radiotélescopes qui existent dans le monde. Son immense portée permettra à l'homme d'étudier systématiquement l'univers entier. Voilà qui, à nos yeux, justifie le choix que nous avons fait de Nançay comme l'un des deux hauts lieux de la science que le lecteur a pu visiter en tête du présent tome. Au cours de cette visite il s'est familiarisé avec les principaux instruments utilisés par les radio-astronomes et a eu un aperçu de la façon dont ils fonctionnent. Mais la radio-astronomie, cette branche nouvelle de l'activité astronomique, ses bases et ses moyens, ses perspectives et les résultats déjà acquis, c'est ici que nous les traiterons en détail.

Avec l'avènement des grands télescopes optiques, Herschel a pu sonder notre univers-île, notre Galaxie; avec l'avènement des grands radiotélescopes, c'est l'univers dans son ensemble qui va pouvoir être sondé. Peut-être pourra-t-on maintenant décider enfin si nous pouvons atteindre ses frontières et s'il se referme sur lui-même. Ces connaissances précises que l'on va obtenir seront fondées sur des études en profondeur faites sur des milliers et des milliers de galaxies et de radiogalaxies. Mais ce ne sont pas seulement ces études à longue portée qui feront faire des progrès décisifs; des connaissances nouvelles fondamentales seront également acquises en ce qui concerne des régions beaucoup plus proches de nous, celles des planètes et de la Lune; plus près encore, c'est même notre propre ionosphère qui sera soumise à un examen détaillé.

ASTRONOMIE

par
Jean HEIDMANN

deux messages : la lumière et l'onde radio

Dans l'état actuel de la technique, la détection des ondes radio-électriques émises par les astres se présente différemment de celle des ondes lumineuses. En astronomie optique, l'œil ou la plaque photographique sont sensibles à des radiations de longueurs d'onde assez variées. Par exemple, l'œil est sensible à toutes les longueurs d'onde comprises entre 0,4 et 0,8 μ; les plaques photographiques ordinaires, à toutes les radiations s'échelonnant de l'ultraviolet au vert. En radio-astronomie, il n'existe pas pour l'instant de récepteur sensible à des plages aussi larges de longueurs d'onde; chacun n'est sensible qu'à une étroite bande de fréquences au voisinage de la fréquence pour laquelle il a été construit. Les radio-électriciens savent bien qu'une antenne, pour fonctionner dans des conditions efficaces, doit être accordée sur une longueur d'onde bien déterminée. De même, un bon récepteur n'est sensible, lui aussi, qu'à une fréquence bien déterminée; il est très sélectif. Si physiquement il n'y a pas de différence essentielle dans la nature des ondes émises par les astres dans le domaine optique ou dans le domaine radio-électrique, puisque l'on passe continûment de l'un à l'autre en changeant la longueur d'onde, les deux domaines sont, au contraire, très séparés technologiquement par leurs méthodes de réception. La vision de l'univers par l'intermédiaire d'un radiotélescope est, au fond, équivalente dans le domaine optique à celle que l'on aurait à travers un filtre coloré ne laissant passer qu'une très étroite bande de longueur d'onde. Une autre différence essentielle avec le domaine optique a une origine purement économique : au foyer d'un miroir de radiotélescope on ne place, en général, qu'une seule antenne réceptrice, suivie d'un seul récepteur, tandis qu'au foyer d'un miroir de télescope optique on place une plaque photographique, c'est-à-dire un très grand nombre de récepteurs optiques, puisque l'on peut considérer que chaque grain de bromure d'argent est un récepteur individuel. Ainsi un radio-astronome aura avec son instrument non seulement une vision de l'univers par l'intermédiaire d'une bande de longueur d'onde très restreinte, mais, en plus, il ne pourra voir à la fois qu'un seul point du ciel.

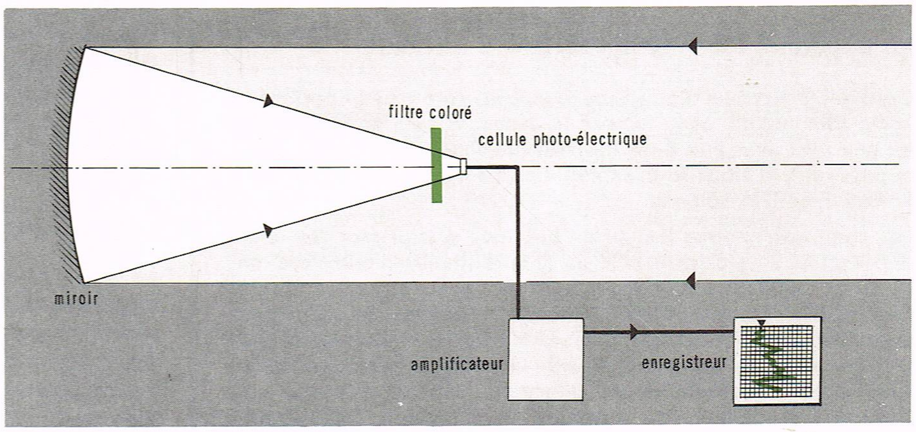
Cela constitue évidemment un gros handicap pour l'observation de l'univers dans un domaine nouveau où il est particulièrement essentiel d'obtenir une large moisson de renseignements très divers. Mais ce n'est pas un handicap fondamental, puisque dans le domaine optique, de plus en plus, se développent des techniques qui sélectionnent des bandes de longueurs d'onde fines, par exemple la raie Hα de l'hydrogène ionisé, et que bien souvent la plaque photographique est remplacée par un petit élément de cellule photo-électrique, beaucoup plus sensible mais n'explorant qu'une toute petite région du ciel à la fois. Lorsque les moyens financiers et la technologie le permettent, on peut essayer de s'affranchir partiellement de ces limitations. C'est ainsi que le grand radiotélescope de Nançay a son foyer équipé de quatre récepteurs fonctionnant sur quatre longueurs d'onde différentes : 6, 11, 21 et 49 cm.

Une radiation est caractérisée par sa longueur d'onde. Les radiations optiques ont des longueurs d'onde d'environ 0,5 micron; les radiations étudiées en radio-astronomie ont des longueurs d'onde allant de 1 cm à 10 m.

La longueur d'onde λ est inversement proportionnelle à la fréquence ν de l'onde : $\lambda = c/\nu$, où c est la vitesse de la lumière; les ondes étudiées en radio-astronomie ont donc des fréquences s'étendant de 30 à 30 000 MHz.

Un astre émet en général des ondes de longueurs très variées, s'étendant depuis le domaine des ondes très courtes, rayons X, jusqu'aux longueurs d'onde très grandes, ondes kilométriques, en passant par l'ultraviolet, le visible, l'infrarouge et les diverses ondes radio.

L'équivalent, dans le domaine de l'optique, d'un radiotélescope serait un miroir au foyer duquel on aurait placé un petit élément de cellule photo-électrique précédé d'un filtre coloré; on ne peut donc explorer à la fois qu'un petit coin du ciel dans une bande étroite de longueurs d'onde.



La courbe donnant l'intensité émise par un astre en fonction de la longueur d'onde, ou de la fréquence, est ce qu'on appelle son spectre.

En mesurant l'intensité reçue sur ces quatre longueurs d'onde, le radio-astronome aura une indication de la forme du spectre des radiations émises par tel ou tel astre. On voit encore ici la situation défavorable de la radio-astronomie par rapport à l'astronomie optique, où pour obtenir le spectre d'un astre il suffit d'en disperser la lumière à l'aide d'un prisme.

La connaissance du spectre des radiations émises par un astre est d'une importance considérable, comme on peut s'en rendre compte en pensant aux progrès énormes et décisifs qui ont été réalisés en astronomie par l'invention du spectroscope. Cela tient au fait que la forme du spectre dépend de façon très intime du processus par lequel un astre émet ses radiations; il est la signature personnelle de l'astre; plus même, il est un message racontant sa vie intime.

les trois émissions radio fondamentales

Trois exemples empruntés au domaine optique vont nous initier aux trois processus fondamentaux d'émission que l'on rencontre dans le domaine radio-astronomique.

Le premier sera fourni par le cas de la lumière émise par un morceau de platine chauffé à 1 500 °C : son spectre est un spectre continu; toutes les longueurs d'onde sont émises; l'intensité, dont le maximum correspond à 2 microns, devient de plus en plus faible lorsqu'on se dirige vers les grandes longueurs ou vers les plus courtes. Ce genre de rayonnement est dû à l'agitation désordonnée des atomes du platine, créée par la température à laquelle on l'a porté; il est appelé *rayonnement thermique*.

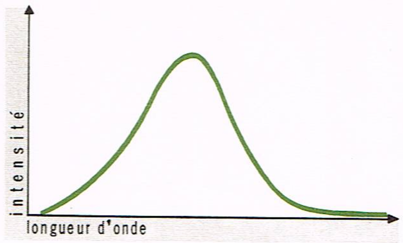
Un autre type de rayonnement est fourni par celui qui est émis par un tube fluorescent; son spectre est réduit à quelques bandes très étroites où l'émission est très intense : les raies. Son origine est très différente : les atomes du gaz dans le tube sont excités par des électrons accélérés entre les anodes du tube; ces atomes, lorsqu'ils se désexcitent et reviennent dans leur état normal, émettent des radiations de longueur d'onde bien déterminée, dépendant de la nature de ces atomes et des états dans lesquels ils ont été excités.

Enfin, un troisième type de rayonnement est fourni par des machines dont il sera question par ailleurs (t. II) : les synchrotrons. Lorsque des électrons très rapides sont astreints à suivre une trajectoire circulaire par le champ magnétique régnant dans le synchrotron, ils émettent des radiations occupant un spectre continu peu large, ayant un maximum à une longueur d'onde d'autant plus courte que l'énergie des électrons est plus grande ou que l'intensité du champ magnétique est plus forte. Ce phénomène a été observé pour la première fois dans le synchrotron de Schenectady, sous forme d'une lueur verdâtre semblant s'échapper du tube où circulaient les électrons.

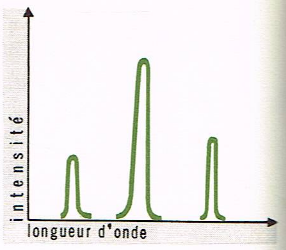
qu'appelle-t-on « température d'antenne » ?

Ces trois processus d'émission jouent un rôle très important en radio-astronomie. Si le lecteur le désire, il peut se familiariser un peu plus avec eux dans les trois paragraphes suivants. Mais, auparavant, il nous faut expliquer un terme que nous aurons à employer très souvent.

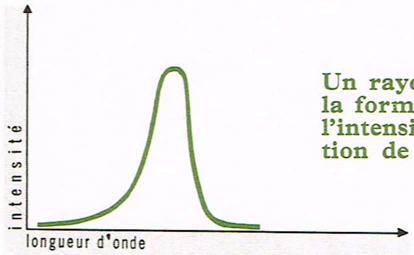
Les radio-astronomes ont pris l'habitude d'exprimer les intensités reçues d'un astre à telle ou telle longueur d'onde non pas en unité habituelle d'énergie, *mais en température*, en degrés absolus Kelvin. Cette façon de faire peut sembler un peu bizarre et déroutante. Elle a des fondements d'origine thermodynamique et technologique qui proviennent de la façon dont les radio-électriciens étalonnent leurs appareils. Si l'on branche une résistance électrique aux bornes d'un récepteur, celui-ci semblera capter une certaine énergie. Cela provient du fait que, la résistance étant à une certaine température, les électrons qu'il contient ont des vitesses désordonnées, résultant de leur agitation thermique. Ainsi se crée un petit courant moyen, qui, pénétrant dans le récepteur, lui fournit une certaine puissance. Les lois de la thermodynamique montrent que la puissance reçue de cette résistance est proportionnelle à la température absolue à laquelle elle est portée. Pour étalonner un récepteur, un radio-électricien remplace donc l'antenne par une résistance ayant la même impédance que l'antenne et portée à une température connue. La température d'un astre mesurée par l'antenne sera alors la température à laquelle il faudrait porter cette résistance pour obtenir sur l'enregistreur une déviation égale à celle qui est fournie par l'astre.



rayonnement thermique



spectre de raies



spectre synchrotron

Un rayonnement est caractérisé par la forme de son spectre, c'est-à-dire l'intensité des radiations en fonction de la longueur d'onde ou de la fréquence.

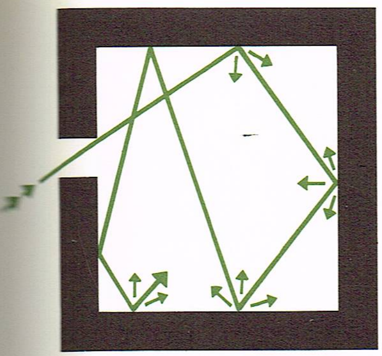
Il existe une autre façon d'opérer pour étalonner un récepteur, mais on ne peut l'appliquer qu'aux petites antennes. Elle consiste à enfermer entièrement l'antenne dans une enceinte ayant les propriétés d'un corps noir, portée à une température uniforme (par exemple : la température ambiante). Cette enceinte peut être constituée, si l'on veut, par une caisse en bois peinte en noir à l'intérieur et recouverte extérieurement d'un grillage empêchant l'échange d'ondes radio avec l'extérieur. Dans ces conditions, un équilibre thermodynamique s'établit et l'antenne se trouve portée à la même température; le courant créé en elle par l'agitation des électrons envoie dans le récepteur une puissance qui est proportionnelle à la température en question. La température d'antenne créée par un astre est alors égale à la température à laquelle il faudrait porter cette enceinte pour obtenir la même déviation sur l'enregistreur.

Qu'on se rappelle enfin qu'une température T, multipliée par la constante de Boltzmann *k*, est égale à une énergie *k*T; et l'on concevra donc qu'il est possible d'exprimer des énergies reçues par des températures équivalentes.

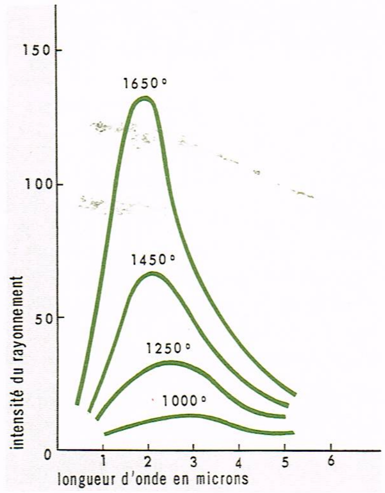
une première émission : le rayonnement thermique

Le cas type du rayonnement thermique est celui du « corps noir » de la physique, enceinte qui absorbe intégralement les radiations de toute longueur d'onde tombant sur elle.

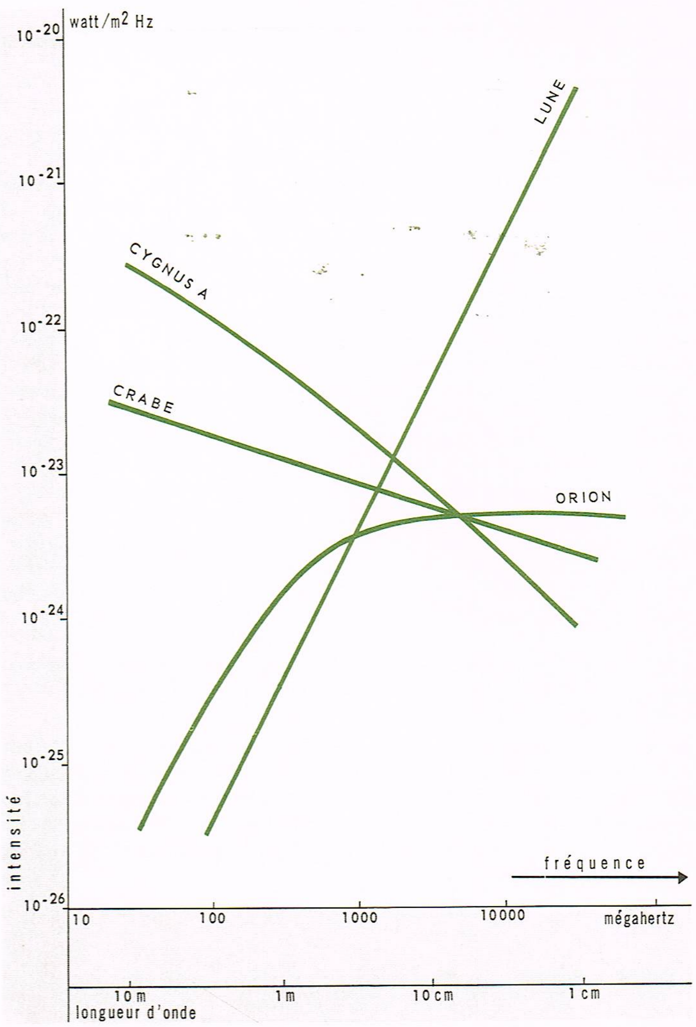
On le schématise souvent par la porte ouverte d'un four : tout rayon lumineux entrant par la porte n'a pratiquement plus aucune chance de sortir du four tant il y subit de réflexions qui



Représentation schématique du corps noir par une petite ouverture percée dans une enceinte.



Spectre du corps noir à plusieurs températures.



Spectre de quelques astres dans le domaine radio-électrique. En ordonnées sont portées les intensités et, en abscisses, les fréquences; la correspondance de celles-ci avec les longueurs d'onde est indiquée en bas.

peu à peu l'amenuisent, pour peu que les parois soient mates (briques, par exemple). Si tout le four est froid, ce qui exclut de placer à l'intérieur une lampe électrique par exemple, l'ouverture paraîtra noire; elle n'émettra pratiquement aucune onde sensible à l'œil. Imaginons maintenant ce four porté à la température de fusion du fer : son ouverture semblera d'un rouge uniforme; si on le chauffait encore plus à blanc, la porte semblerait blanche. On voit ainsi que le rayonnement émis aura une composition variant avec la température; il sera d'autant plus intense que la température est plus forte. La loi exacte est donnée par la célèbre formule de Planck, formule où n'intervient qu'un paramètre physique : la température du corps. En radio-astronomie, où l'on a affaire à de grandes longueurs d'onde, cette formule se simplifie en loi de Rayleigh-Jeans : l'énergie émise est proportionnelle à la température du corps et au carré de la fréquence. La température d'antenne que l'on mesurait en plaçant un tel corps devant le radiotélescope est indépendante de la fréquence; c'est celle du corps noir.

Pour un corps non parfaitement absorbant, l'énergie émise est celle du corps noir multipliée par le pouvoir absorbant qui, lui, peut être une fonction de fréquence. C'est ainsi qu'un miroir parfaitement réfléchissant n'émet rien, car il n'absorbe rien.

Pour un gaz, le problème est plus compliqué, du fait de sa transparence. Il faut considérer chaque élément du gaz avec son pouvoir absorbant propre et additionner les rayonnements produits le long des trajets des rayons.

La situation sera différente selon l'épaisseur du gaz. Si la couche de gaz est très épaisse, il se comporte comme un corps opaque et la température d'antenne que l'on observe en pointant vers ce gaz est tout simplement la température du gaz. Si, par contre, le gaz est peu épais, la température d'antenne sera plus petite et variera en proportion de l'épaisseur du gaz, ce qui est bien intuitif.

Cette notion d'épaisseur de gaz est très importante; elle dépend non seulement de l'épaisseur vraie, mais aussi de la densité et de la transparence du gaz; comme cette dernière varie avec la longueur d'onde en général, l'épaisseur dépend elle aussi de la longueur d'onde; c'est pourquoi on la nomme « épaisseur optique », mais le terme « épaisseur électromagnétique » serait plus approprié.

On peut représenter ces résultats graphiquement (en portant en abscisses la fréquence et en ordonnées l'intensité observée).

Le cas du corps noir est représenté par la droite de pente 2, marquée Lune. Le cas d'un gaz est illustré par la courbe marquée Orion; sur la gauche, elle a une pente 2 : le gaz de la nébuleuse est épais, tandis que sur la droite la pente diminue : le gaz devient mince à haute fréquence, car le coefficient d'absorption diminue pour un gaz ionisé lorsque la fréquence augmente.

type de rayonnement	exemple optique	exemple radio-électrique
thermique	lampe à incandescence	la Lune
de raie	tube fluorescent	nuages d'hydrogène atomique neutre
synchrotron	machine synchrotron	nébuleuse du Crabe

La nature profonde du rayonnement thermique provient de l'équilibre statistique qui s'établit à l'intérieur du corps entre les mouvements des électrons et les énergies des photons qui sont émis et absorbés lors des transitions des électrons d'un état à un autre.

une deuxième émission : sur 21 cm

« La raie de 21 cm est due à une structure hyperfine de l'état fondamental de l'atome d'hydrogène neutre », peut-on lire dans les traités de physique. Voyons, en langage imagé, ce que cela veut dire. On peut se représenter l'atome d'hydrogène comme formé d'un électron circulant autour d'un proton; les orbites qu'il peut décrire sont déterminées par les conditions de la mécanique quantique. Dans l'état normal, l'électron circule sur l'orbite la plus proche du proton. En lui fournissant de l'énergie, à l'aide d'un photon, on peut le faire passer sur une orbite plus éloignée. Inversement, si l'électron redescend sur l'orbite fondamentale, un photon bien défini sera émis : la différence d'énergie des deux états est égale à l'énergie du photon et à ce photon correspond une onde lumineuse dont la longueur est bien définie. C'est ainsi qu'on explique les « raies », c'est-à-dire les radiations de longueur d'onde définie, que les atomes émettent. Mais un raffinement, un hyperraffinement pourrait-on dire, doit être apporté : en réalité, le proton et l'électron doivent être considérés comme deux petits aimants, et selon que, dans leur mouvement relatif, ces petits aimants seront orientés dans le même sens ou dans des sens opposés, leur énergie sera différente. **C'est lors du passage de l'état parallèle à l'état antiparallèle que la « raie de 21 cm » est émise.**

La différence d'énergie qui existe entre ces deux états possibles est extrêmement faible et correspond donc à une longueur d'onde relativement très grande.

Lorsqu'un radiotélescope détecte le rayonnement sur 21 cm venant d'une masse d'hydrogène atomique, la raie en sera élargie par les mouvements de turbulence ou de rotation des éléments de ce gaz par l'effet Doppler-Fizeau, **et c'est justement dans la possibilité de détermination de vitesses radiales en radio-astronomie que réside le grand intérêt des mesures de 21 cm.**

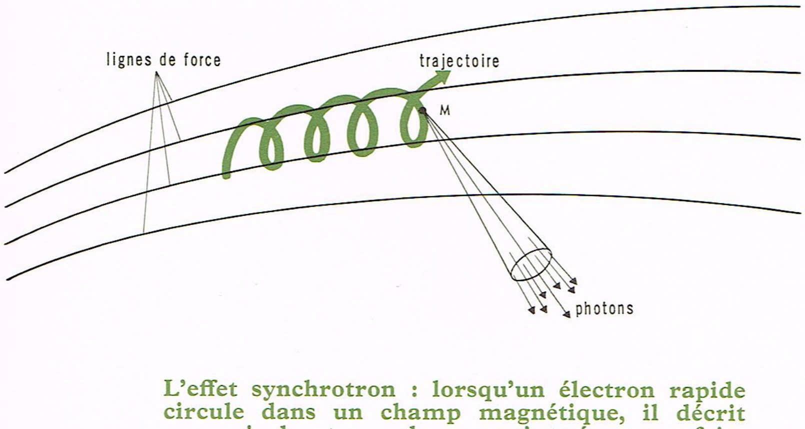
Toute notre Galaxie devient accessible à l'étude cinématique, alors que cette étude est rendue pratiquement impossible dans le domaine optique par l'énorme absorption qui règne dans le plan de la Voie lactée. La forme du profil de la raie, c'est-à-dire l'intensité mesurée en fonction de la longueur d'onde, permettra de calculer le nombre d'atomes d'hydrogène qui se déplacent par rapport à nous à une certaine vitesse selon la ligne de visée du radiotélescope. Comme dans le cas du rayonnement thermique, on peut établir une formule reliant la température d'antenne mesurée à la profondeur optique et à la température des électrons. Tout comme en optique, il sera d'ailleurs possible d'observer la raie de 21 cm en émission ou en absorption.

Dans le domaine des ondes radio, il existe d'autres raies d'origine atomique, ou moléculaire; la moins difficile à détecter serait l'homologue de la raie de 21 cm pour le deutérium, qui se situe à 91 cm. Des essais de détection ont été infructueux jusqu'à maintenant; ils permettent seulement de dire que la proportion de deutérium dans l'hydrogène interstellaire est inférieure au millième. Une autre raie qui serait éventuellement observable serait celle du radical OH à 18 cm. Les résultats négatifs sont cependant intéressants par les limites supérieures qu'ils donnent à l'abondance de certains éléments dans les milieux intersidéraux.

Des mesures difficiles ont également été entreprises pour essayer de détecter l'effet de champs magnétiques (effet Zeemann) sur la raie de 21 cm. La mesure d'un tel effet permettrait d'estimer la valeur du champ magnétique régnant dans le milieu interstellaire. Malheureusement, on n'a pu jusqu'à maintenant obtenir qu'une limite supérieure du champ magnétique de l'ordre de 10^{-5} gauss.

une troisième radiation : le rayonnement « synchrotron »

Une source importante de rayonnement dans le domaine radioastronomique est le rayonnement dit « synchrotron ». Si un électron se déplace dans un champ magnétique, il décrit une spirale et émet des ondes à la fréquence avec laquelle il décrit les spires comme le ferait un circuit oscillant. Si cependant l'électron a une très haute énergie, un effet de contraction relativiste fait que son émission ne s'opérera que dans un cône étroit centré autour de la vitesse de l'électron.



L'effet synchrotron : lorsqu'un électron rapide circule dans un champ magnétique, il décrit une spirale et, en chaque point, émet un faisceau de radiations compris dans un petit cône autour de sa vitesse en ce point.

De plus, cette distribution en faisceau a pour effet d'élargir la bande dans laquelle les radiations sont émises; le spectre du rayonnement synchrotron sera continu, quoique peu large. Si l'on connaît la répartition des électrons en énergie dans un milieu où règne un certain champ magnétique, il sera possible de calculer l'intensité du rayonnement émis par ces électrons en fonction de la fréquence. Ici apparaît le lien qui relie la radioastronomie à la physique des rayons cosmiques. Les rayons cosmiques ont une répartition en énergie connue, et il est naturel de supposer que les électrons relativistes ont également la même répartition. On obtient en définitive pour le spectre émis une courbe descendante.

C'est effectivement ce qui est observé pour des radio-sources aussi différentes physiquement que la radiogalaxie géante du Cygne A ou le reste de la supernova du Crabe.

Il existe bien d'autres types d'émission de rayonnement radio dont le rôle est cependant plus secondaire. Le type le plus important est celui des « oscillations de plasma », nuage de particules positives mêlé à un nuage de particules négatives. Si l'on essaie de séparer le nuage positif du nuage négatif, une violente attraction tend à les faire revenir l'un vers l'autre; ainsi naissent des « oscillations de plasma ». Le processus par lequel ces oscillations peuvent émettre des ondes radio est encore assez obscur, mais on pense que ce processus est efficace dans certains phénomènes se déroulant dans la couronne solaire, par exemple : en ondes métriques, les températures d'antenne peuvent y atteindre des valeurs de plusieurs dizaines de milliards de degrés (il faut se rappeler ici le caractère un peu artificiel de la définition des températures d'antenne : c'est la température que devrait avoir un corps noir pour émettre autant que ces oscillations de plasma. Mais, physiquement, il n'y a aucun rapport entre l'émission thermique du corps noir et une oscillation de plasma, comme nous l'avons vu).

quelques radio-sources particulièrement proches

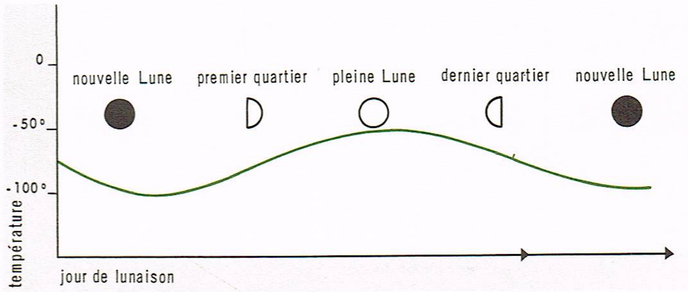
Après ces préliminaires un peu théoriques, nous pouvons commencer notre visite de l'univers radio-astronomique. Quittant la Terre et ses parasites dus aux orages, aux industries, quittant ses émissions radio intelligentes — ou du moins organisées —, nous traversons les lignes de force du champ magnétique terrestre, siège lui aussi d'émissions radio de très grandes longueurs d'onde, pour atteindre le premier corps céleste, laissant de côté les émissions des nombreux satellites qui maintenant entourent notre globe.

L'émission radio de la Lune est d'origine purement thermique. Sur ondes centimétriques, on trouve que la température d'antenne est d'environ 200° absolus (c'est-à-dire 73 °C au-dessous de zéro) avec des variations au cours de la lunaison d'amplitude maximale 30°. Ces variations sont décalées par rapport à la lunaison d'environ trois jours, du fait du « volant thermique » présenté par le sol lunaire. Sur ondes décimétriques, les résultats sont similaires, mais les variations avec la lunaison sont, en revanche, pratiquement indécelables; cela provient du fait que l'épaisseur « optique » des roches lunaires augmente si la longueur d'onde augmente; les mesures décimétriques se réfèrent donc à des couches plus profondes du sol lunaire où, par volant thermique, les variations sont naturellement beaucoup plus faibles. Des informations très intéressantes peuvent ainsi être obtenues sur la nature du sol lunaire par son émission thermique.

Nous éloignant encore de la Terre, nous atteignons les planètes.

Mars d'abord : son émission semble, elle aussi, être purement thermique et, à cause de la petitesse de la planète, très difficile à détecter; sur ondes centimétriques, la température trouvée est de l'ordre de 200° absolus, en accord avec la température mesurée pour le rayonnement infrarouge. Pour Vénus, sur ondes centimétriques également, une température excédant 600° absolus (soit plus de 300 °C) a été mesurée, tandis que les mesures infrarouges donnent une température moitié moindre. Cela est dû au fait que les ondes radio proviennent du sol de Vénus, tandis que les mesures infrarouges proviennent des couches supérieures de l'atmosphère très épaisse de la planète.

Jupiter, sur ondes centimétriques, donne une température de l'ordre de 150° absolus, en accord, ici, avec les températures mesurées par le rayonnement infrarouge. En ondes décimétriques, cependant, la température mesurée devient beaucoup plus forte, atteignant 100 000° pour une longueur d'onde de 1 mètre; le rayonnement n'est donc ici nullement d'origine thermique. D'autre part, des mesures de forme de la planète sur 30 cm de longueur d'onde ont montré qu'elle était un ellipsoïde très aplani dont le diamètre équatorial est de l'ordre de trois fois le diamètre de la planète optique. Il semble qu'il faille expliquer l'origine de cette émission par des électrons relativistes circulant dans un champ magnétique intense qui entourerait la planète Jupiter. En ondes décamétriques, le comportement de Jupiter est encore plus curieux. On observe des sursauts brefs d'émission durant quelques secondes et très intenses, dont l'intensité peut atteindre 1 000 fois celle du Soleil vue de la Terre. Ils agitent les enregistrements pendant une heure ou deux, puis disparaissent. La fréquence d'apparition de ces sursauts est liée à la période de rotation, comme si leur origine était fixée en un point particulier du sol, point particulier qui d'ailleurs peut changer de place d'année en année. A plus grande fréquence, ces sursauts sont complètement invisibles, peut-être à cause d'un effet d'écran introduit par l'ionosphère de la planète. S'agit-il de gigantesques orages dans l'atmosphère de Jupiter ou de perturbations apportées par des éjections de gaz venant de l'océan agité de la planète, on ne le sait encore.



La température de la Lune, mesurée en ondes centimétriques, est d'environ 73 °C au-dessous de zéro ; elle varie légèrement avec la lunaison, avec un retard de quelques jours par rapport aux phases de la Lune.

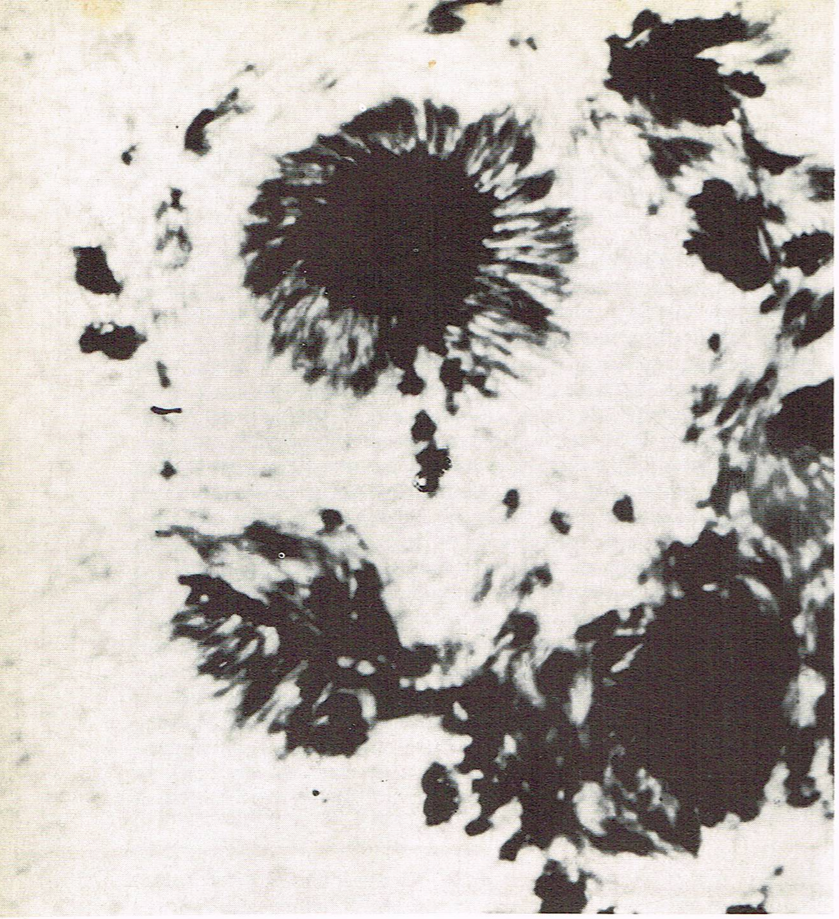
les rayonnements typiques du Soleil

Revenons vers le centre du système solaire. Croisant auparavant quelques comètes dont l'émission radio n'a jusqu'à maintenant jamais pu être décelée, saluant au passage quelque sonde planétaire envoyée par l'homme, nous arrivons au Soleil.

Les différents types d'émissions solaires sont fort complexes, leur complexité dépendant de la longueur d'onde à laquelle ils ont lieu, de leur mode de variation dans le temps... On peut cependant tenter de les classer selon les situations physiques dans lesquelles, pense-t-on, ils ont leur origine. On peut distinguer trois types principaux de situations.

Premièrement, celle du Soleil calme, c'est-à-dire du Soleil des minimums d'activité solaire; l'émission est purement thermique; c'est celle de l'atmosphère solaire considérée dans son ensemble : photosphère, chromosphère, couronne. Mais au phénomène d'émission thermique viennent se superposer des phénomènes de propagation dus au fait que, d'une part, cette atmosphère est ionisée et que, d'autre part, la densité des électrons décroît régulièrement lorsqu'on s'élève au-dessus de la surface du Soleil. Dans un milieu ionisé, une onde ne pourra se propager que si sa fréquence est supérieure à une certaine fréquence critique augmentant avec la densité des électrons; ainsi des ondes décamétriques ne pourront provenir que de régions du Soleil situées à au moins un rayon solaire de la photosphère. Des ondes décimétriques, elles, pourront provenir de la haute chromosphère et, bien évidemment, des ondes optiques pourront provenir de la photosphère. On conçoit que l'étude radio du Soleil calme permette de sonder l'atmosphère solaire par tranches successives, selon la fréquence utilisée. Tel est l'objet des **Années internationales du Soleil calme.**

Un deuxième type de situation est celui qui est présenté par ce que les opticiens appellent un centre d'activité caractérisé par la présence d'une plage faculaire et de taches solaires. L'émission radio de ces centres d'activité est aussi une émission purement thermique, mais d'intensité plus forte, provenant des condensations coronales; les techniques radio permettent ainsi d'étudier ces condensations non seulement lorsqu'elles se présentent en profil sur le bord du disque solaire, mais aussi en plein centre du disque solaire. De véritables cartes solaires peuvent être ainsi tracées, montrant un Soleil plus grand que

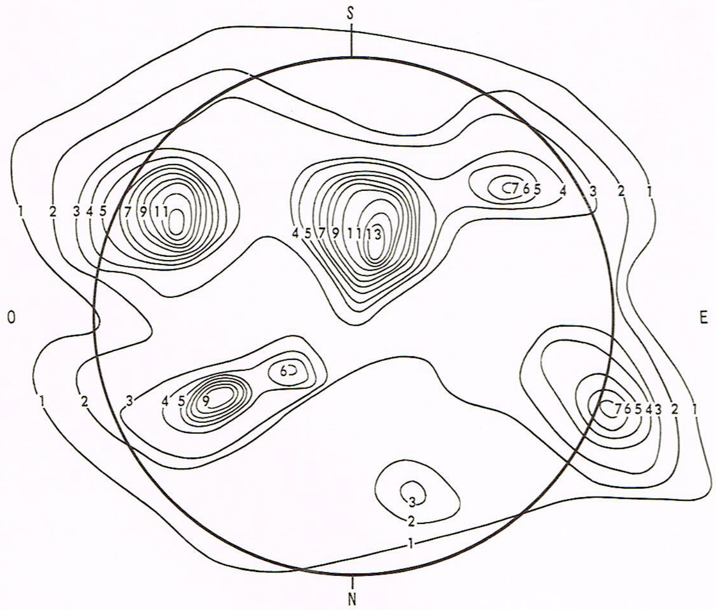


Taches solaires photographiées à l'aide d'un télescope emporté par un ballon à 24 000 m d'altitude, hors des couches troublées de l'atmosphère.

le Soleil optique. Ces émissions radio sont bien visibles sur ondes décimétriques; l'intensité globale du Soleil varie d'environ 50 p. 100 avec une période de 27 jours.

Une troisième situation est celle qui résulte des éruptions chromosphériques. Ces phénomènes solaires explosifs donnent lieu à des émissions radio de caractère non thermique. Deux types d'éjections peuvent être créés par une éruption chromosphérique : premièrement, celle d'un nuage ionisé qui, lorsqu'il est projeté à travers l'atmosphère solaire, la perturbe et donne naissance, par des oscillations de plasma ou par ondes de choc, à des émissions radio de fréquence de plus en plus basse au fur et à mesure que le nuage s'élève dans l'atmosphère solaire. La vitesse de ces nuages a pu ainsi être mesurée et trouvée égale à environ 1 000 km par seconde. Peut-être s'agit-il là des nuages qui, au bout de quelques jours, viendront perturber le champ magnétique terrestre. Une éruption chromosphérique peut aussi éjecter des particules de très haute énergie à la vitesse de la lumière; lorsque ces particules relativistes traversent l'atmosphère solaire, elles créent, par effet synchrotron dans son champ magnétique, des émissions radio observables. Les plus intenses de ces émissions correspondent aussi à l'éjection de rayons cosmiques par le Soleil.

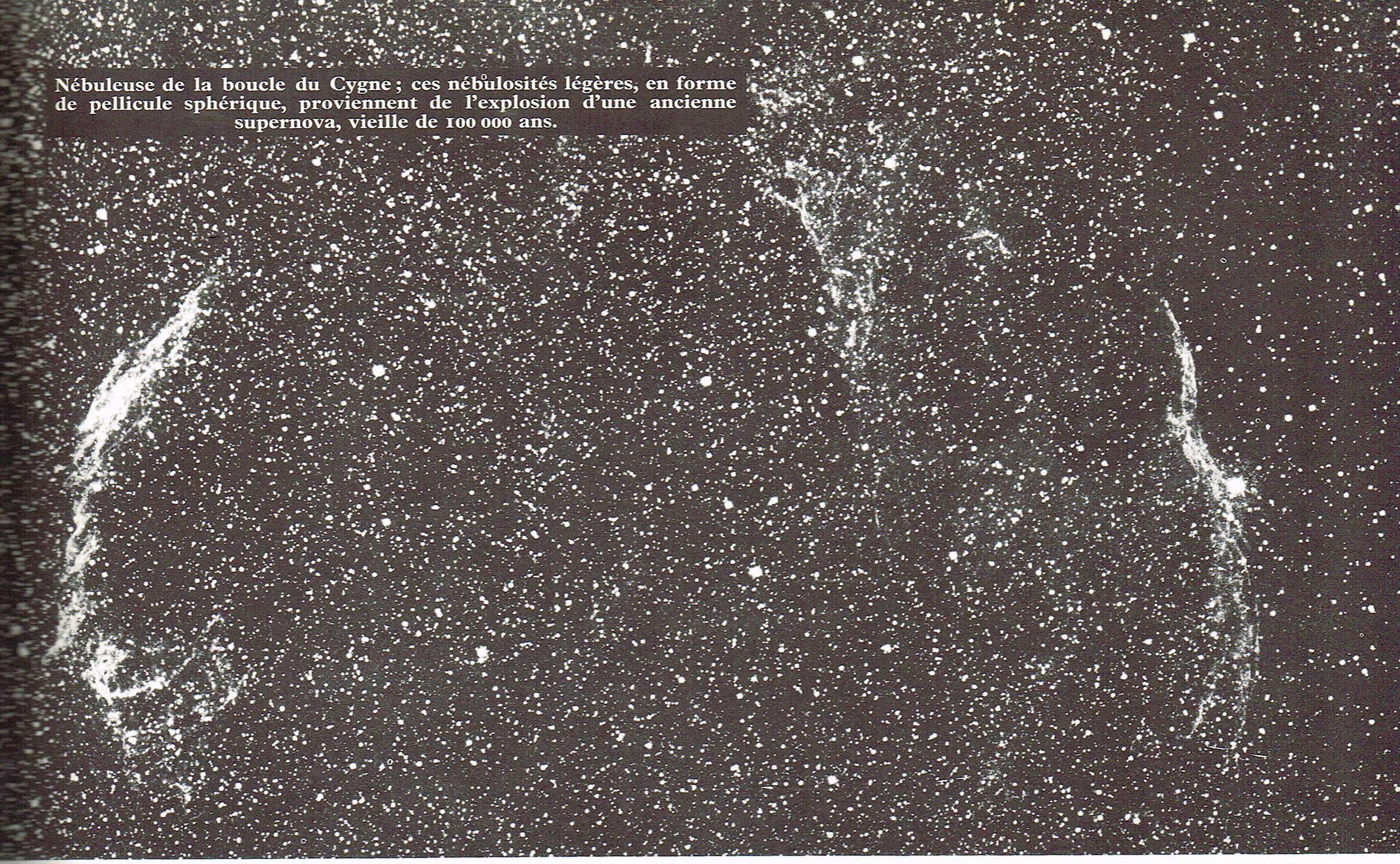
Il existe encore un type d'émission dit « orage de bruit ». Il s'agit de sursauts très brefs sur ondes métriques; leur origine — mal connue — semble se situer très haut dans la couronne. Les grandes lignes de l'activité radio-électrique du Soleil se sont déroulées partiellement au cours de ce dernier cycle, les phénomènes exacts par lesquels les ondes sont émises sont encore incompris et, comme nous l'avons déjà dit, le Soleil est véritablement un laboratoire d'études de réactions entre plasma. Grâce à la radio-astronomie, l'étude d'un milieu ionisé à haute température est possible; de plus, les liaisons que l'on peut établir entre phénomènes solaires et phénomènes géophysiques sont maintenant beaucoup plus fournies et pourront servir à prévoir ceux-ci.





Vision simultanée du Soleil en optique et en radio. *Ci-contre*, photographie en raie Hα, le 3 décembre 1957; *ci-dessus*, carte en longueur d'onde de 21 cm le même jour. Sur la photographie du Soleil a été superposée une photographie de la couronne.

Phot. U. S. I. S.



Nébuleuse de la boucle du Cygne; ces nébulosités légères, en forme de pellicule sphérique, proviennent de l'explosion d'une ancienne supernova, vieille de 100 000 ans.

les nébuleuses gazeuses

Il nous faut maintenant entreprendre un voyage 200 000 fois plus long pour atteindre les plus proches étoiles. La plupart d'entre elles sont certainement le siège d'activités comparables à celles du Soleil, mais leur distance est telle que jusqu'à maintenant rien de sûr n'a pu être détecté.

Mais d'autres objets vont se révéler à nous et certains d'entre eux d'aspect tellement nouveau, tellement insolite, qu'ils donnent à la radio-astronomie son caractère révolutionnaire si attrayant. Le voyage est long, non seulement en distance puisqu'il ne se terminera qu'au « bout » de l'univers, mais aussi en temps, car il nous ramènera des milliards d'années en arrière, aux « débuts » de l'univers.

Il faut atteindre les régions H II pour trouver des astres radio-astronomiquement puissants. Les régions H II sont des nébuleuses gazeuses portées à une température de l'ordre de 10 000 °C par une ou deux étoiles très chaudes situées en leur sein. On observera donc le rayonnement thermique d'un gaz ionisé.

Le spectre d'une telle région H II aura l'aspect caractéristique que présente Orion; à grande fréquence, le gaz est peu épais et l'intensité est indépendante de la fréquence, alors qu'à basse fréquence le gaz est épais et l'intensité augmente avec la fréquence.

De nombreuses radiosources autres que celle d'Orion ont été identifiées dans des régions H II : nébuleuses Oméga, América, de la Rosette, etc. Ces régions étant en général situées dans le plan de la Galaxie, beaucoup d'entre elles sont cachées à notre vue par les nuages de matière absorbante de la Voie lactée; aussi, de nombreuses sources thermiques ne peuvent-elles être mises en correspondance optique avec des régions H II. Ces nébuleuses ont, en général, un diamètre angulaire faible, inférieur à 1 degré et il a fallu faire des mesures interférométriques pour pouvoir préciser leur forme. Ayant mesuré les dimensions de la nébuleuse (1 minute d'arc sur 3, pour Orion), on peut préciser la valeur de la densité électronique en son sein et obtenir des indications sur l'état physique qui y règne.

les supernovae : bel exemple d'effet synchrotron

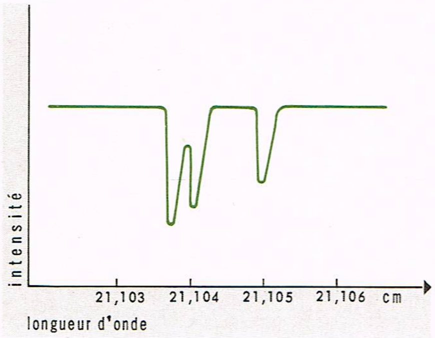
Une autre classe importante de radiosources galactiques est celle des restes de supernovae. Les plus importantes sont celle de Kepler, apparue en 1604, celle de Tycho Brahé, en 1572, la boucle du Cygne, la nébuleuse du Crabe, apparue en 1054, et enfin Cassiopée A, la plus intense radiosource du ciel. Le spectre radio-électrique de ces objets indique que l'émission est due à l'effet synchrotron. Ces restes de supernova sont vus à des stades d'évolution différents; la boucle du Cygne est probablement un exemple très ancien de supernovae; par contre, la nébuleuse du Crabe est observée dans un état beaucoup plus proche de sa naissance, puisqu'elle n'a que neuf cents ans. Son aspect optique est d'ailleurs beaucoup plus concentré et l'objet radio lui correspond très bien en position et en forme.

Phot. Texereau (Observatoire de Haute-Provence, télescope de Schmidt).

les nuages d'hydrogène

Comme pour les régions H II, de nombreuses radiosources galactiques non thermiques doivent nous être cachées optiquement par les nuages de matière obscure qui parsèment la Voie lactée.

Pour pouvoir observer d'autres objets galactiques en radioastronomie, il faut avoir recours à la raie de 21 cm. On a pu mettre en évidence des nuages individuels d'hydrogène atomique (régions H I) en étudiant les lignes d'absorption 21 cm dans le spectre de radiosources lointaines. Imaginons que l'on pointe un radiotélescope vers une radiosource puissante, à spectre radio continu. S'il existe sur la ligne de visée des nuages d'hydrogène atomique, ceux-ci créeront de l'absorption dans le spectre continu de la source; on reconnaît, par exemple, sur le spectre de Cassiopée A, la présence de trois raies d'absorption créées par trois nuages, un dans le bras local dit d'Orion, deux dans le bras plus lointain de Persée. On peut ainsi dire que Cassiopée A est éloignée d'au moins 10 000 années de lumière. Cette méthode permet d'estimer les distances des radiosources galactiques. De plus, la largeur intrinsèque de chacune des raies permet d'avoir une idée de la dispersion des vitesses des atomes d'hydrogène dans chacun des nuages et de mesurer ainsi leur température. Il se confirme que cette température est basse, inférieure à 100° absolus et qu'il s'agit bien de régions H I.



Dans le spectre de certaines radiosources, on observe dans le voisinage des longueurs d'onde de 21 cm des raies d'absorption dues à des nuages d'hydrogène atomique neutre interposés entre la source et nous.

Une application inattendue de ces méthodes d'observation a été faite à la détermination de la distance de la Terre au Soleil. Voici comment : la radiosource du Crabe est à peu près située dans le plan de l'écliptique; elle présente aussi dans son spectre quelques raies d'absorption dues à des nuages d'hydrogène interstellaires. En mesurant avec une grande précision la fréquence à laquelle se placent ces raies à six mois d'intervalle, il est possible de mesurer la vitesse de la Terre autour du Soleil, et la troisième loi de Kepler permet d'en déduire le rayon de l'orbite terrestre, valeur qui, jusqu'à l'application des méthodes de radar-astronomie à la planète Vénus, était très mal connue.

l'architecture complexe de la Galaxie

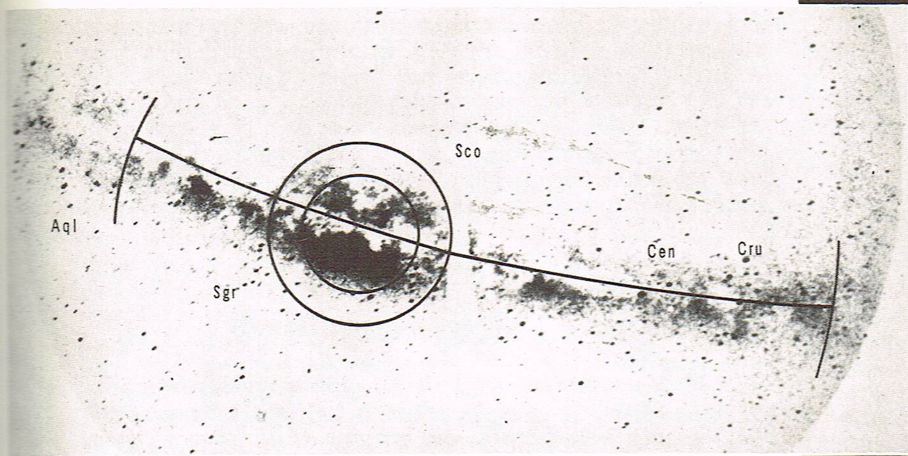
Ici s'arrête le monde radio-astronomique des petits objets; entendons-nous, par petits objets nous voulons dire les astres qui ont des dimensions inférieures à une centaine d'années de lumière. L'enveloppe de la supernova de Cassiopée, par exemple, a une dizaine d'années de lumière. Des régions H II typiques ont plusieurs dizaines ou jusqu'à une centaine d'années de lumière de diamètre. Mais il s'agit quand même bien de petits objets par rapport aux autres radiosources; à commencer par les galaxies, et notamment la nôtre, qui a un diamètre de 100 000 années de lumière.

Le lecteur trouvera peut-être que nous avons été bien vite en chemin, mais notre intention est de freiner l'allure, maintenant que nous atteignons le domaine des galaxies; et cela est un peu justifié par le fait que le grand miroir de Nançay est destiné aux études de ce domaine principalement. Notre Galaxie est une agglomération d'une centaine de milliards d'étoiles occupant un volume ellipsoïdal très aplati. On s'en fait une image relativement fidèle en regardant la photographie de la Nébuleuse N. G. C. 891. C'est un peu à dessein que nous choisissons une galaxie vue exactement par la tranche, car c'est l'aspect que présente la nôtre vue de notre Terre, puisque nous sommes dans le plan de notre Galaxie comme le montre cette photographie infrarouge à grand champ de la Voie lactée. Sur ces deux photographies, qui ont beaucoup de points communs, nous remarquons la présence d'un gros noyau central composé d'une immense agglomération d'étoiles; ce noyau est prolongé de chaque côté par un disque aplati, composé lui aussi d'étoiles et divisé en deux par une bande fine de matière absorbante. Voilà à peu près à quoi se bornait notre connaissance de la structure galactique avant qu'on l'étudie sur 21 cm. Il était pratiquement impossible d'étudier optiquement sa structure étant donné que nous sommes dans son plan et que ce plan est encombré de nuages absorbants. L'aspect d'autres galaxies vues de face incitait à penser évidemment que la nôtre possédait aussi un système de bras spiraux. En mesurant la répartition dans l'espace des étoiles intrinsèquement les plus brillantes (étoiles O et B) il a été possible de mettre en évidence l'existence de trois tronçons de bras spiraux au voisinage de notre Soleil : un dans lequel nous sommes pratiquement plongés, un autre plus près du centre de la Galaxie et un autre à l'extérieur. Des mesures de vitesse radiale d'étoiles ont également permis de mettre en évidence et de mesurer la rotation de notre Galaxie. On a découvert que les étoiles de notre voisinage se mouvaient dans une direction perpendiculaire à la direction du Sagittaire, où gît le centre de la Galaxie; cette rotation présente, en outre, l'aspect d'une rotation différentielle, c'est-à-dire que les étoiles plus proches du centre ont une vitesse plus grande que les étoiles plus éloignées du centre, exactement comme dans le cas du mouvement des planètes autour du Soleil; autrement dit, les étoiles de notre voisinage ont des mouvements circulaires autour de la masse centrale de notre Galaxie suivant la troisième loi de Kepler. Cette découverte permettait d'obtenir la première estimation de la masse totale de la Galaxie et d'obtenir une première idée du nombre de ces étoiles; soit 100 milliards. Mais des progrès décisifs dans la connaissance de la structure de notre Galaxie partent de cette dernière décennie, et cela grâce aux mesures effectuées dans la raie de 21 cm de l'hydrogène.

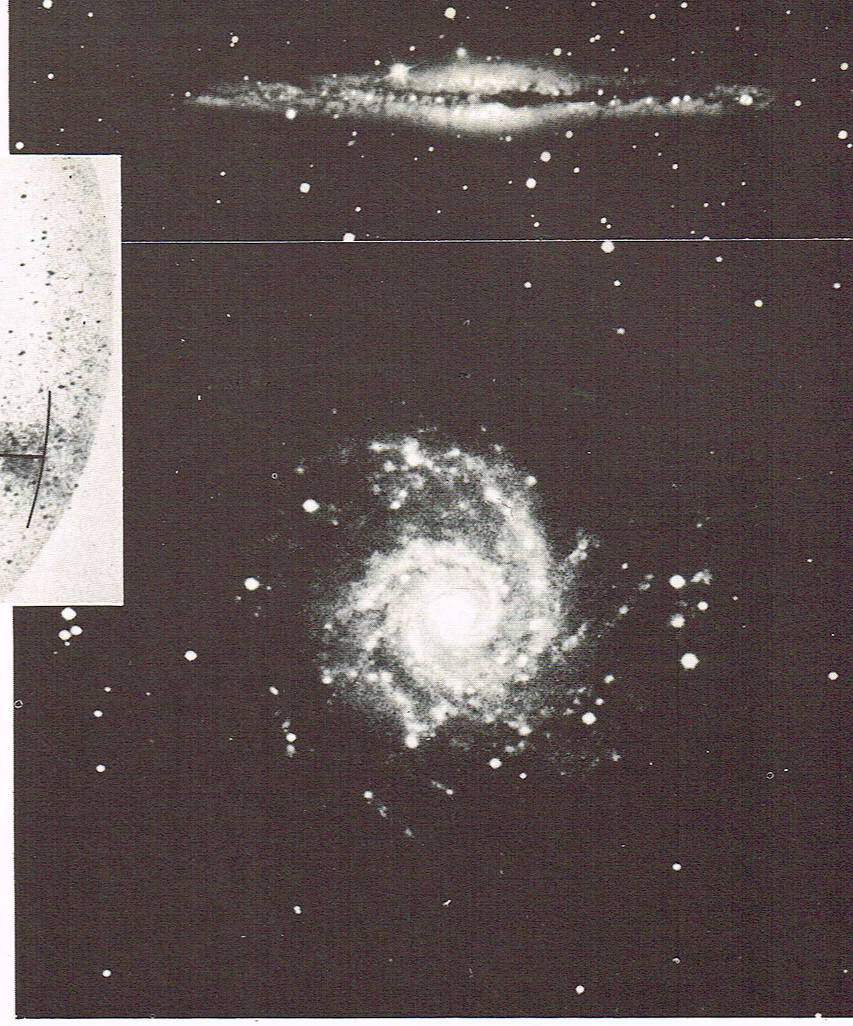
ces bras spiraux qui nous entourent

La possibilité de détecter la raie 21 cm dans l'hydrogène interstellaire a été prévue à l'avance par Van de Hulst, jeune étudiant hollandais, vers 1943. Dès la fin des hostilités, trois observatoires, situés dans trois pays différents, se mettaient à l'ouvrage et réussissaient à la détecter presque simultanément. Bien sûr, on ne pouvait espérer déterminer directement que la quantité d'hydrogène mise en jeu et sa vitesse radiale; mais une heureuse particularité cinématique, propre à notre Galaxie, a permis d'en déduire la distance au Soleil des objets émetteurs.

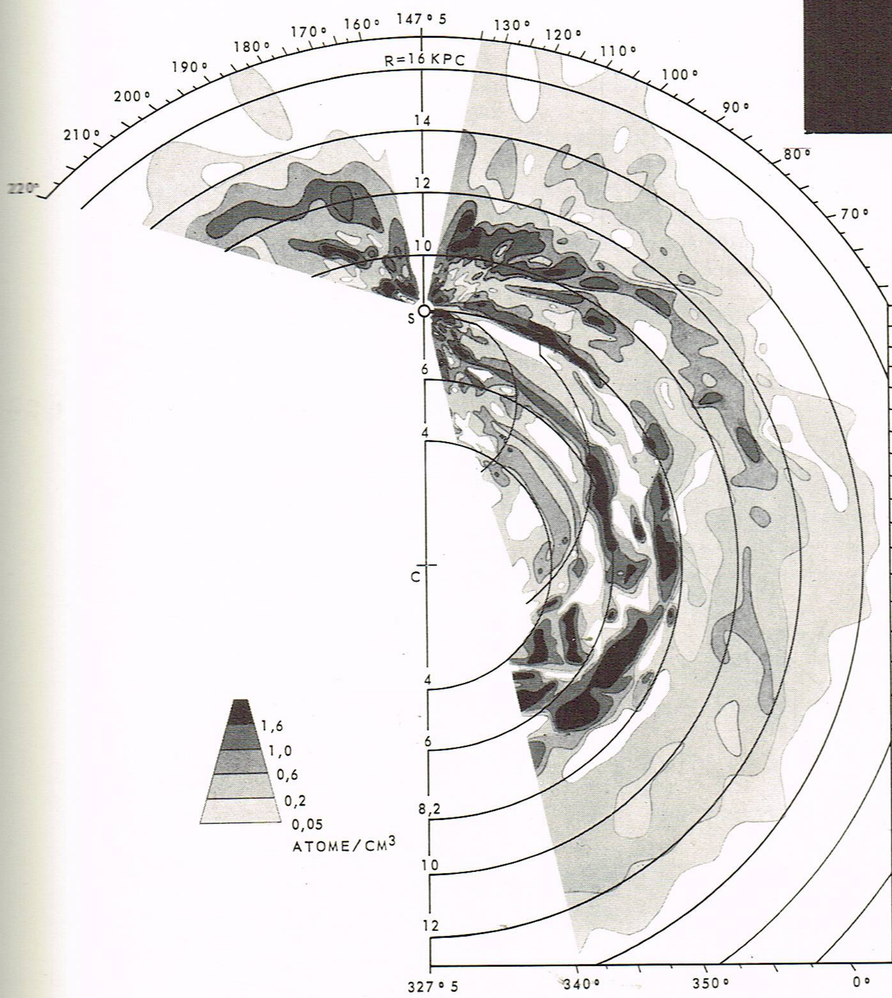
Une carte générale de la distribution de l'hydrogène atomique dans la Galaxie peut être dressée; la densité de l'hydrogène varie de 1/10 à 1 et plus atome d'hydrogène par centimètre cube. On y distingue trois bras principaux, qui sont nommés,



La Voie lactée, photographiée en infrarouge avec un objectif à grand champ dans la direction du centre, ressemble effectivement à N. G. C. 891. *(Doc. « Astrophysical Journal ».)*

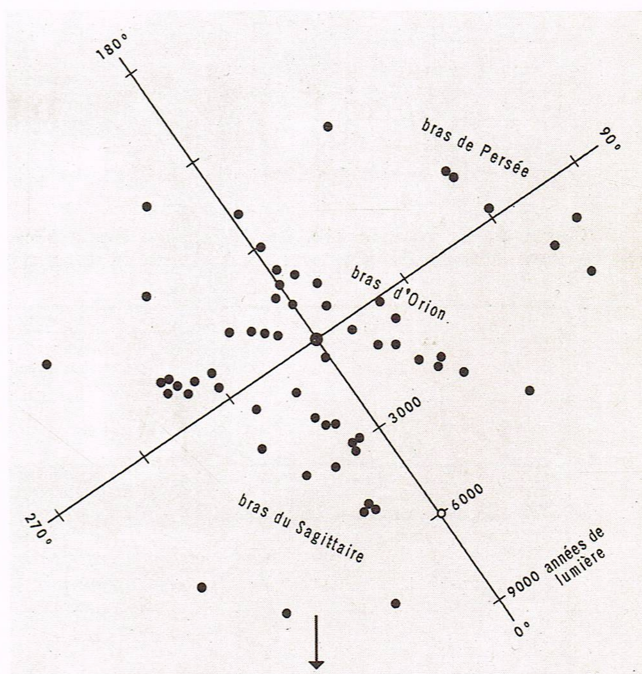


Deux galaxies spirales : N. G. C. 891, vue de profil, et M 74, vue de face.



Répartition de l'hydrogène atomique neutre dans le plan de notre Galaxie selon les mesures des astronomes hollandais, sur 21 cm ; notre Soleil est en S. *(Doc. B. A. N.)*

Chaque point indique la position dans le plan de notre Galaxie d'une étoile O ou B, le gros point représentant notre Soleil et la flèche la direction du centre galactique. Ces fortes étoiles se groupent selon trois tronçons de bras spiraux.



Phot. Observatoire Lick (université de Californie).

de l'intérieur vers l'extérieur, les bras du Sagittaire, d'Orion et de Persée, le Soleil se trouvant sur le bord intérieur du bras d'Orion. L'angle des spirales avec le rayon vecteur est faible, de l'ordre de 5 degrés; la Galaxie est donc assez enroulée. Le Soleil se déplaçant vers la droite, les bras sont « à la traîne » et la rotation différentielle tend à les enrouler de plus en plus. A tel point, d'ailleurs, qu'on est obligé d'admettre que les bras sont des structures de caractère tout à fait transitoire.

Dans la direction du centre galactique, les observations sont pratiquement impossibles à débrouiller, car les vitesses radiales sont toutes faibles et l'absorption est importante. Cependant, en observant la radiosource non thermique Sagittaire A, qui est située en plein centre de notre Galaxie, comme nous le verrons, on a observé dans le spectre continu de cette source une forte raie d'absorption à la fréquence de 21 cm. Le recoupement avec d'autres mesures conduit à attribuer cette absorption à un bras d'hydrogène qui serait situé à 10 000 années de lumière du centre galactique. Le fait tout à fait remarquable est que ce bras n'est pas en mouvement circulaire, mais qu'il a une vitesse radiale d'expansion de 50 km par seconde. L'hypothèse du mouvement circulaire n'est donc qu'une hypothèse approchée, et des travaux considérables sont en cours depuis plusieurs années pour en tenir compte dans le dépouillement des observations. Dans la région du Soleil, cette vitesse d'expansion ne serait que de quelques kilomètres par seconde. On est conduit à admettre que chaque année une masse d'hydrogène atomique égale à environ une masse solaire est éjectée du centre de notre Galaxie.

un espace interstellaire rempli d'hydrogène et où naissent les étoiles

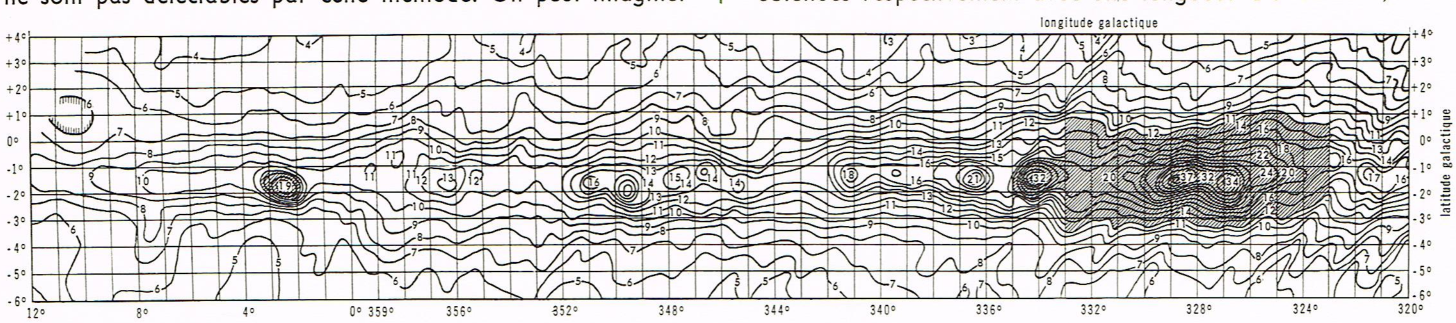
La masse totale de l'hydrogène atomique détectée est de l'ordre de quelques pour cent de la masse de la Galaxie; c'est une faible quantité, mais il ne faut pas oublier qu'il ne s'agit que de l'hydrogène atomique; l'hydrogène ionisé et l'hydrogène moléculaire ne sont pas détectables par cette méthode. On peut imaginer que l'hydrogène des bras est réparti en une multitude de nuages tels que ceux qui sont observés en absorption. *C'est dans ces nuages que prendraient naissance les étoiles nouvelles. L'hydrogène des bras serait alors progressivement transformé en étoiles, puis ces étoiles, passant par le stade explosif de nova ou de supernova, réinjecteraient dans la Galaxie de nouvelles quantités de gaz; ainsi pourraient peut-être s'expliquer le mouvement de circulation de l'hydrogène, la formation et la dissolution continue des bras spiraux.*

un vide interstellaire parcouru par des champs magnétiques

Dans la dynamique de tels processus doit intervenir la science nouvelle constituée par la magnéto-hydrodynamique. Cette spécialité est, rappelons-le, une branche de la mécanique qui étudie le mouvement de gaz fortement ionisés plongés dans des champs magnétiques. De nombreuses indications semblent prouver que la Galaxie possède un champ magnétique. D'autre part, même s'il est peu ionisé, le gaz interstellaire, de par ses dimensions énormes, est doué d'une conductivité électrique pratiquement infinie. Il en résulte que les lignes de force magnétique sont liées aux masses gazeuses et que la dynamique de celles-ci ne peut se dissocier de l'étude des mouvements de ces lignes de force, c'est-à-dire qu'il y a de puissants effets électromagnétiques tels que, par exemple, l'accélération des rayons cosmiques. Il faut donc parler de magnéto-hydrodynamique de la Galaxie, faisant intervenir la gravitation des étoiles, l'écoulement gazeux de gaz magnétisé, l'énergie de turbulence magnétique et l'énergie des rayons cosmiques.

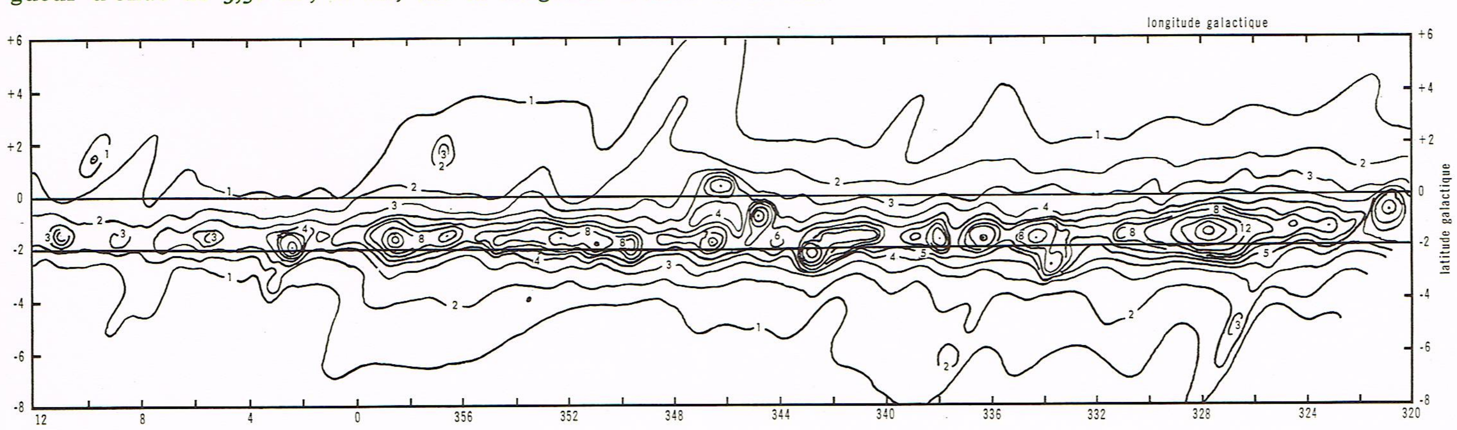
disque et halo galactiques

D'autres informations expérimentales peuvent être obtenues sur la structure générale de la Galaxie par l'étude du continuum radio. Voici ci-dessous côte à côte deux cartes célestes obtenues respectivement avec une longueur d'onde de 3,50 m



Carte de la Voie lactée, de l'Aigle au Sagittaire. *En haut,* **sur la longueur d'onde de 3,50 m;** *en bas,* **sur la longueur d'onde de 22 cm.**

(D'après Hill, Slee et Mills, en haut; et G. Westerhout, en bas.)



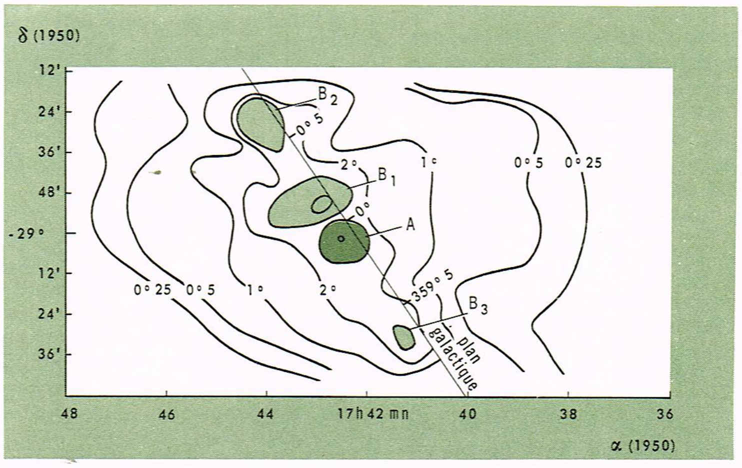
et une longueur d'onde de 22 cm; elles sont à la même échelle et les antennes ont les mêmes lobes (exploitent le même champ). Les coordonnées sont galactiques, le plan galactique étant au milieu de chaque figure et le centre de la galaxie vers la droite à 327° de longitude. On voit de nombreuses sources isolées, que l'on peut reconnaître d'une carte à l'autre; certaines sont identifiables à des nébulosités optiques. On remarque aussi une crête de forte intensité qui suit l'équateur galactique, s'intensifie lorsqu'on se rapproche du centre galactique. Cette crête se superpose à un fond continu général beaucoup plus intense en ondes métriques qu'en ondes décimétriques, où il est pratiquement invisible. Ce fond continu, qui a une température de plusieurs milliers de degrés en ondes métriques et seulement de quelques degrés en ondes décimétriques, n'est certainement pas d'origine thermique. On l'attribue à l'émission synchrotron d'électrons relativistes qui se déplaceraient dans un halo galactique général, englobant notre Galaxie, constitué d'un gaz ionisé de très faible densité et de très haute température. Quant à la crête galactique, on l'attribue au rayonnement d'un grand nombre de sources, thermiques et non thermiques (par exemple régions H II et restes de supernovae), et à un supplément de radiation non thermique qui serait dû à des électrons relativistes circulant dans un champ magnétique relativement intense (de l'ordre peut-être de 10^{-5} gauss) qui existerait à l'intérieur des bras spiraux de notre Galaxie.

au centre de la Galaxie

Avant de quitter notre Galaxie, il nous faut fixer notre attention sur son centre, dont la radio-astronomie a révélé toute une série de propriétés extraordinaires. Cette région semble jouer un rôle tout à fait particulier, non seulement au point de vue morphologique, mais aussi, comme nous le verrons à propos des radiogalaxies, au point de vue génétique. Le centre de certaines galaxies est le siège de phénomènes absolument exceptionnels, dont il y a quelques années on n'avait aucune idée et qui, actuellement, semblent bien mystérieux. En optique, le centre galactique nous est irrémédiablement caché. En revanche, par les ondes radio, le centre de notre Galaxie révèle sa structure, une structure très complexe. Sur la longueur d'onde de 350 cm, elle se présente sous la forme d'une source double occupant une zone d'environ 1 degré sur 2. Sa nature est probablement d'origine non thermique, puisqu'elle est très intense en ondes métriques, et sa dualité serait expliquée par l'absorption créée en son milieu par une autre source, thermique, occupant le centre de notre Galaxie.

Cette source est bien observée avec une intensité forte sur 22 cm. Des cartes beaucoup plus détaillées obtenues sur des longueurs d'onde centimétriques ont révélé la structure compliquée de cette source centrale. Des précisions ont pu être apportées par des mesures interférométriques. Il s'avère que cette source est elle-même composée d'une part d'une source circulaire d'un diamètre de 3′ d'arc, très intense et de nature probablement thermique (source A), et, d'autre part, d'une source allongée selon le plan galactique et de forme très tourmentée (B_1 B_2 B_3). Sa nature n'est pas encore connue. Etant donné la distance du centre de la Galaxie, la petite source intense A a un diamètre de l'ordre de 30 années de lumière. Cette structure est à rapprocher de celle du noyau d'aspect quasi stellaire qui est observé au centre de notre grande voisine, la nébuleuse d'Andromède. Il se présente sous la forme d'une « étoile » très floue, d'un diamètre de quelques secondes. Si l'on considère la distance de cette nébuleuse, cela correspond à un diamètre de 40 années de lumière. Récemment un spectre de cet amas d'étoiles a été pris grâce à la caméra électronique et a révélé une rotation très rapide atteignant 100 km/s au bord. Les lois de la gravitation conduisent alors à une densité stellaire 10 000 fois supérieure à celle qui existe au voisinage de notre Soleil.

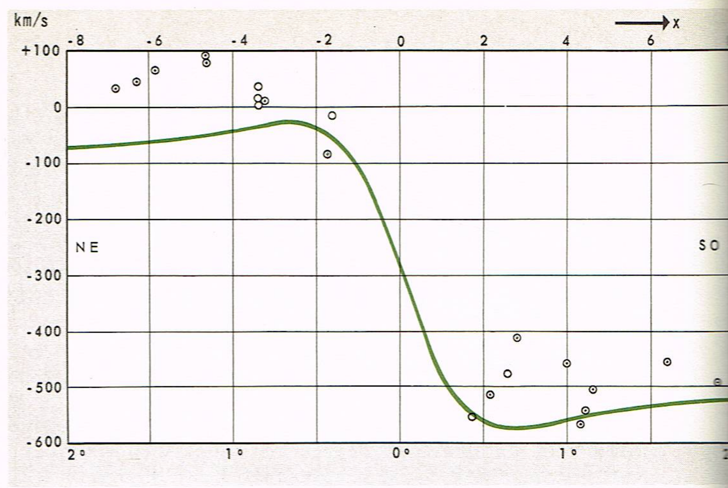
Il semble, toujours d'après les mesures sur 21 cm, que le centre de notre Galaxie soit occupé par un anneau plat d'hydrogène de 500 années de lumière de diamètre. A l'intérieur de cet anneau se trouverait un disque d'hydrogène dont la densité augmenterait très rapidement vers le centre. Cette structure d'hydrogène est en rotation très rapide, 250 km/s, et les lois de la gravitation confirment les grandes densités stellaires trouvées par les autres observations. **Il semble donc que le centre de notre Galaxie soit occupé par une concentration stellaire très importante entourée d'une région complexe;** cette région complexe contient un gaz abondant qui, en plus des vitesses et de l'ionisation observées, est le siège d'émissions synchrotron puissantes; il contient donc des champs magnétiques et des électrons relativistes. C'est peut-être dans cette région que les bras spiraux prennent naissance.



Carte des régions centrales de notre Galaxie, en ondes centimétriques. Les courbes représentent les lignes d'égales températures mesurées. *(D'après Drake.)*



Région centrale de l'amas de galaxies Coma.



Courbe de rotation de la nébuleuse d'Andromède. Les vitesses radiales y sont données en kilomètres-seconde en fonction de la distance au centre de la Galaxie le long du grand axe.

Photographie de la nébuleuse d'Andromède. Le cercle représente le lobe d'un miroir de 25 m et l'ellipse allongée celui du grand miroir de Nançay, tous deux sur ondes de 21 cm.

l'architecture globale du cosmos

Si nous voulons continuer ce voyage dans le cosmos, il nous faut maintenant quitter la Galaxie; son diamètre, rappelons-le, ne dépasse pas 100 000 années de lumière, alors que c'est à 2 millions d'années de lumière qu'il nous faut nous éloigner maintenant.

Notre Galaxie, Andromède et quelques autres de taille plus faible constituent ce qu'on appelle le « groupe local de galaxies ». Pour atteindre d'autres galaxies dans l'espace, il faut voyager encore plus loin. Le prochain groupe se trouve dans la Grande Ourse, à 8 millions d'années de lumière. L'espace est ainsi peuplé d'innombrables galaxies; les plus lointaines enregistrées sont pour l'instant à 5 milliards d'années de lumière. Mais n'anticipons pas. Le grand radiotélescope de Nançay est destiné surtout à la détection de la raie de 21 cm dans les galaxies extérieures, et sa portée pour cette raie ne dépassera pas 100 ou 200 millions d'années de lumière. Le rayonnement de 21 cm est quand même un rayonnement assez faible et, jusqu'à maintenant, avec les télescopes mis en service de par le monde dont les diamètres ne dépassent pas 25 m, seules nos plus proches voisines, une dizaine en tout, ont pu être mesurées. Avec sa surface de 7 000 m², le télescope de Nançay pourra pénétrer beaucoup plus loin dans l'espace, car il sera beaucoup plus sensible. Il aura aussi le très net avantage d'avoir un pouvoir de résolution beaucoup plus grand.

notre voisine, Andromède

La figure de la page ci-contre représente une photographie de notre voisine Andromède; on distingue aussi deux satellites de cette galaxie. Le cercle représente le champ de visée, le lobe, d'un télescope de 25 m de diamètre, et la tache allongée selon la direction nord-sud, c'est le lobe du miroir de Nancay (3′ 30″ de large sur 20′ [1] de haut). La forme allongée du lobe résulte du fait que la surface réceptrice de Nançay est allongée dans le sens est-ouest. Déjà avec des télescopes de 25 m des résultats très intéressants ont été obtenus en ce qui concerne nos voisines.

(1) Le symbole ′ désigne des minutes d'arc, ″ des secondes d'arc, alors que mn et s désignent des minutes et des secondes de temps (unités dont on confond trop souvent les significations alors qu'il s'agit de notions bien différentes).

On sait, par exemple, que la région centrale tourne d'un bloc; la vitesse atteint un maximum de près de 300 km/s et, lorsqu'on continue de s'éloigner vers les régions extérieures de la galaxie, la vitesse décroît suivant les lois de Kepler. La vitesse d'ensemble de la galaxie par rapport à notre système a pu être elle aussi déduite avec une grande exactitude; elle vaut 300 km/s. La forme de la courbe de rotation a permis d'établir un modèle de distribution de masse à l'intérieur d'Andromède et d'estimer sa masse totale à quatre fois celle de notre Galaxie. Enfin, l'intensité totale de la raie de 21 cm provenant de cette nébuleuse a permis de mesurer la proportion d'hydrogène qu'elle contient à l'état atomique; on trouve 1 p. 100 seulement de la masse totale. Mais les miroirs de 25 m ne permettent pas, avec un lobe aussi large par rapport à l'image de la galaxie, d'étudier en détail la structure spirale, alors qu'avec le lobe de Nançay il sera possible de faire une véritable carte d'Andromède et de voir comment se dispose l'hydrogène atomique par rapport à l'image optique; de plus, il va être possible d'étudier très en détail la dynamique de cette galaxie, par exemple de voir s'il y existe aussi des mouvements d'expansion.

pourquoi étudier les galaxies sur 21 cm

On peut estimer que le grand miroir de Nançay pourra étudier environ mille galaxies, en fait les mille galaxies les plus brillantes de la voûte céleste. Le domaine encore très mal connu de la dynamique des galaxies va donc pouvoir s'ouvrir largement aux investigations; il sera même possible de balayer toute la zone d'espace qui nous est cachée par les matières absorbantes de notre Voie lactée et d'y découvrir quelques nouvelles voisines de notre système. Cette découverte serait la bienvenue, car seuls Andromède et le Triangle présentent un aussi grand diamètre angulaire et peuvent être fouillés avec autant de détails, du moins dans notre hémisphère. Dans l'hémisphère Sud, deux galaxies occupent des diamètres angulaires beaucoup plus importants; ce sont les deux nuages de Magellan, satellites très proches gravitant autour de notre Galaxie. A part ces objets situés d'une façon exceptionnelle, on peut compter en étudier quelques dizaines d'autres, qui seront, au point de vue expérimental, dans la situation de la nébuleuse d'Andro-



mède par rapport au miroir de 25 m. D'excellentes courbes de rotation pourront être obtenues. Ensuite, il y aura un lot d'une à deux centaines de galaxies dont la largeur sera de l'ordre de deux ou trois lobes; pour celles-ci, il sera seulement possible de faire une distinction entre la gauche et la droite, et d'obtenir un embryon de courbe de rotation. Enfin, en ce qui concerne la majorité des mille galaxies accessibles, l'image sera de l'ordre de grandeur du lobe; la radiation de 21 cm qui en proviendra sera alors vue globalement. Cependant, de la forme de cette raie, de son profil, il sera possible de déduire des informations dynamiques. Imaginons, par exemple, une galaxie vue exactement de face; d'après ce que l'on sait actuellement, les mouvements perpendiculaires au plan d'une galaxie doivent être faibles; donc la raie d'une telle galaxie devra être très fine, alors que tout élargissement observé de la raie apportera des indications sur les mouvements perpendiculaires au plan de la galaxie. Pour une galaxie vue de profil, la raie occupera toute une bande de fréquences, dont les valeurs extrêmes permettront de déduire les vitesses de rotation maximales existant dans cette galaxie. Pour les galaxies dites « irrégulières », on observera peut-être plusieurs raies simultanées, correspondant à des mouvements de convection à grande échelle.

De toute façon, la quantité totale d'hydrogène pourra être mesurée. On pourra la comparer aux caractéristiques optiques principales : magnitude, type. La relation de l'hydrogène neutre avec les types de population stellaire jeune ou vieille pourra être étudiée sur place, ce qui apportera des informations sur la création des étoiles et les phénomènes d'ionisation; par exemple, jusqu'à maintenant il a été impossible de mesurer une elliptique géante, la plus proche, en effet, se trouvant dans l'amas de la Vierge, à près de 100 millions d'années de lumière. Certains pensent que ces galaxies sont des galaxies vieilles, dont les étoiles n'évoluent pratiquement plus. Elles seraient vides de tout gaz, et pourtant l'une d'entre elles, comme nous le verrons par la suite, est douée d'une activité radio-électrique remarquable.

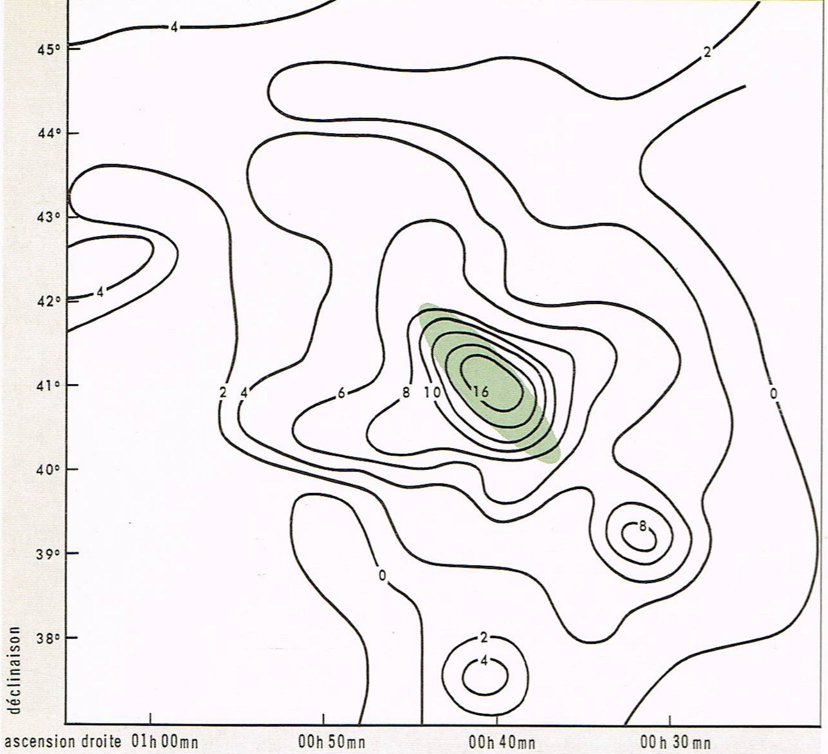
les groupes de galaxies

Un grand nombre de vitesses radiales de galaxies vont pouvoir aussi être mesurées et la dynamique des groupes de galaxies sera abordée. Souvent, les galaxies apparaissent par paires ou par trois, par exemple notre Galaxie et ses deux satellites — les nuages de Magellan —, Andromède et ses deux satellites. Une dizaine de ces groupes proches seront à la portée des radiotélescopes, et même un grand amas, celui de la Vierge. Le problème de la dynamique de ces groupes n'a pu encore être attaqué de front avec de grands moyens et les questions en suspens sont nombreuses. Ainsi, pour certains groupes on observe que les vitesses radiales relatives de leurs membres sont énormes et l'on est conduit à des alternatives opposées. Si ces galaxies ont effectivement la masse qu'on leur suppose d'après d'autres observations, les systèmes ne sont pas stables gravitationnellement et doivent se disperser. Etant donné les vitesses relatives et les dimensions physiques du système, on peut calculer le temps de vie de ce système et l'on trouve qu'il est très faible, quelquefois inférieur à 1 milliard d'années. Il en résulte nécessairement que les galaxies de ces groupes ont un âge inférieur à cette valeur. **Le problème de la genèse des galaxies dans l'espace se pose alors.** Ou bien on est conduit à admettre que les galaxies ont des masses beaucoup plus grandes que celles qui ont été mesurées par ailleurs; ou bien encore que ces groupes contiennent une grande quantité de gaz pour les lier gravitationnellement.

les amas de galaxies

La dynamique des grands amas de galaxies tels que l'amas de la Vierge sera elle aussi accessible. La nature vraiment particulière de ces amas vaut que l'on s'y arrête un peu. L'amas de Coma est un exemple typique de grand amas compact; il est situé à environ 300 millions d'années de lumière de nous et occupe dans le ciel une portion circulaire de plusieurs degrés de diamètre. Il groupe peut-être 10 000 galaxies, la densité de celles-ci augmentant régulièrement et fortement lorsque l'on s'approche de son centre. Par suite de l'expansion de l'univers, il s'éloigne de nous à la vitesse de 7 000 km/s, la raie d'hydrogène doit donc être déjà décalée de 35 MHz vers les basses fréquences. La spectroscopie optique a mesuré avec beaucoup de peine les vitesses radiales de quelques dizaines de ces galaxies. Ces vitesses radiales indiquent des dispersions de plusieurs milliers de kilomètres à la seconde.

Lorsque le lobe du grand miroir sera pointé vers l'amas de Coma, il sera parsemé de nombreuses galaxies, et l'on aura d'un coup leur vitesse moyenne d'ensemble. En explorant l'amas tranche par tranche, son état dynamique sera précisé : mouvements radiaux ou circulaires des composants, expansion ou rotation d'ensemble, équilibre statistique atteint ou non, relation avec la forme. Ces questions ont une importance considérable pour l'étude à grande échelle de temps et d'espace de l'évolution de l'univers. On s'en rend compte en songeant qu'il faut plusieurs milliards d'années pour qu'une galaxie de l'amas de Coma le traverse. Si l'on admet que les interactions s'y produisent actuellement par gravitation entre galaxies, on conçoit que l'état dynamique de l'amas soit un reflet fidèle du stade physique, peut-être très différent, amas de gaz par exemple, par lequel il a dû passer lors de sa formation, ce qui donnerait ainsi en



Carte de la région de la nébuleuse d'Andromède en longueur d'onde de 70 cm. *(Doc. « Nature ».)*

quelque sorte une empreinte fossile de l'état passé de l'univers. Selon les réponses qui seront apportées par les mesures expérimentales, on peut même être amené à des révisions révolutionnaires, telle cette suggestion récente selon laquelle l'amas de Coma aurait plus de 10^{15} ans. Sur les cartes de comptage de galaxies à la surface du ciel on peut dénombrer une vingtaine d'amas aussi importants que celui de Coma et même des amas d'amas. **C'est un domaine nouveau d'investigation qui s'ouvre avec ses immenses perspectives.**

Récemment, on a découvert des chaînes de galaxies où l'on voit quatre ou cinq nébuleuses disposées en ligne. Ces chaînes incitent à penser que des galaxies se forment actuellement, car ces chapelets ne peuvent, selon les idées actuelles, avoir qu'une existence très brève, de par leur forme très instable. L'état primordial d'une galaxie, le plus naturel à concevoir, est celui d'un immense nuage d'hydrogène froid, donc neutre, contenant cent fois plus de ce gaz que les galaxies normales. S'il existe de ces « pré-galaxies », le grand miroir devrait les détecter à des distances dix fois plus grandes que pour les galaxies déjà formées.

Le lecteur va penser que nous anticipons vraiment trop. Peut-être, mais cela a du moins le mérite de lui montrer que l'avenir de la radio-astronomie est riche de promesses. De plus, nous tenons à préciser que ces anticipations sont fondées sur des suppositions d'un caractère tout à fait classique, et nous pensons que l'étude des galaxies extérieures par la raie de 21 cm constituera le domaine le moins révolutionnaire de la radio-astronomie extra-galactique.

Nous pouvons maintenant porter notre attention sur l'étude des galaxies extérieures en continuum radio. D'emblée il nous faut faire une distinction entre deux classes très différentes de galaxies : les galaxies dites « normales » et les radiogalaxies « anormales », ces dernières constituant une des découvertes majeures de la radio-astronomie.

les galaxies normales

Par galaxie normale, on entend une galaxie ressemblant à la nôtre ou à celle d'Andromède, s'il s'agit d'une géante, ou au satellite elliptique de cette dernière, s'il s'agit d'une naine. En bref, une galaxie normale est actuellement considérée comme une galaxie n'ayant pas dans le domaine radio de propriétés plus spéciales que celles de notre Galaxie.

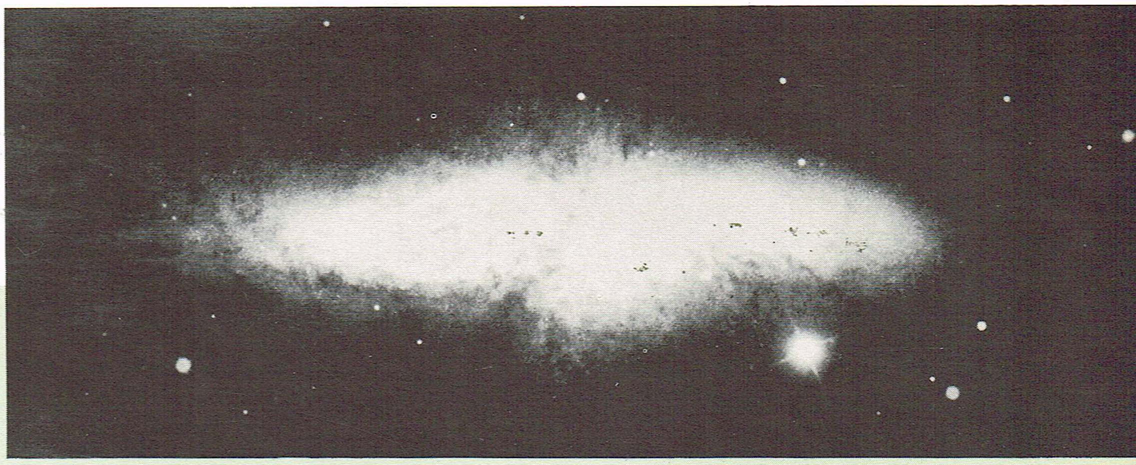
Précisons tout de suite que ces déterminations sont déjà difficiles, puisque les températures mesurées n'excèdent pas quelques dizaines de degrés au-dessus du fond continu du ciel, qui atteint 100 ou 200 °K. On est un peu dans la situation de l'astronome « optique » qui voudrait photographier une faible nébuleuse par une nuit de pleine lune. L'observation des galaxies normales par le continuum radio présentera donc une difficulté du même ordre que celle qui se rencontre avec la raie de 21 cm de l'hydrogène. Considérons la carte ci-contre de la région d'Andromède. L'ellipse au centre de la figure représente le contour optique de la galaxie. Les pics qu'on observe sont probablement dus à des sources extra-galactiques beaucoup plus lointaines. Cette carte confirme d'ailleurs la structure obtenue pour notre Galaxie.

A part les nébuleuses d'Andromède et du Triangle, seules une vingtaine d'autres galaxies proches ont pu être détectées en ondes métriques. L'intensité radio émise par ces galaxies permet de leur attribuer une magnitude radio, comme, en optique, l'intensité lumineuse permet de définir une magnitude optique. La comparaison des magnitudes radio et optique pour ces galaxies normales montre que l'une et l'autre varient parallèlement, c'est-à-dire que la différence magnitude radio moins magnitude optique, que l'on appelle « indice radio », est une constante, ce qui veut simplement signifier en définitive que l'intensité radio pour les galaxies normales est proportionnelle à l'intensité lumineuse. Malheureusement, les galaxies les plus brillantes de notre voisinage sont en général les spirales et, jusqu'à maintenant, aucune elliptique ou irrégulière n'a pu être mesurée. Cela montre l'intérêt qu'il y aura à posséder un radiotélescope plus puissant.

quelques radiogalaxies particulièrement intéressantes

Nous quittons maintenant le domaine des galaxies normales, mais, avant d'atteindre celui des galaxies violemment anormales, telles que celles que l'on a longtemps appelées les « galaxies en collision », nous aurons à fixer notre attention sur toute une suite de types intermédiaires, tous aussi intéressants les uns que les autres, mais tous aussi différents les uns des autres que cela est possible et tous également incompris pour l'instant. Comme aucune explication générale du phénomène de la radiogalaxie anormale n'a encore été élaborée, nous devrons nous contenter de cette description phénoménologique et nous adopterons l'ordre des intensités intrinsèques croissantes de l'intensité radio. La seule chose dont on soit à peu près sûr pour l'instant est que ces émissions radio anormales sont dues à l'effet synchrotron d'électrons relativistes dans des champs magnétiques.

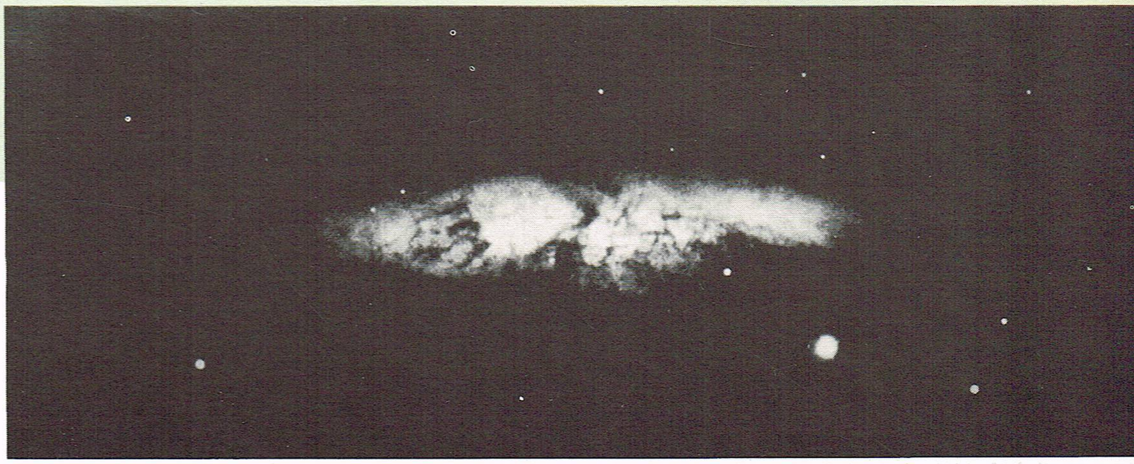
Jusqu'à maintenant les quelques dizaines de radiogalaxies identifiées optiquement n'ont pas permis d'établir une relation entre le type radio et le type optique. Des galaxies d'apparence optique tout à fait normale sont anormales au point de vue radio, mais il faut dire que si Messier 87 était situé à grande distance, son jet optique serait invisible. Des galaxies simples optiquement correspondent à des galaxies doubles au point de vue radio, et, inversement, des galaxies doubles optiquement correspondent à des galaxies simples au point de vue radio, bien que ce ne soit pas le cas le plus fréquent. Des galaxies doubles optiquement et qui présentent de grands signes d'interaction réciproque ne sont pas émetteurs radio. Certaines radiogalaxies ont un spectre optique tout à fait ordinaire; réciproquement, des galaxies à spectre hautement excité ne sont pas des émetteurs radio puissants. Aucun indice ne vient encore pour l'instant éclairer l'énigme des radiogalaxies; tout ce que l'on peut dire c'est qu'elles sont, en général, doubles ou quadruples et que leur spectre radio est « synchrotron ».



pose de 30 mn

La galaxie irrégulière Messier 82.

pose de 5 mn



Messier 82 : une explosion récente ?

Messier 82 est une de nos proches voisines, puisqu'elle appartient à l'amas de la Grande Ourse, à 8 millions d'années de lumière. Elle fait partie d'une classe très particulière de galaxies irrégulières, le type « irrégulier II » : sa structure optique ne semble suivre aucun plan, à part sa forme allongée. Mais elle a ceci de très particulier qu'il y a à son intérieur des traces d'absorption extrêmement marquées. Sur la photographie moins exposée, l'image est barrée en tout sens de traînées en zigzag de matière noire, certaines régions étant complètement oblitérées. On pense que cette absorption est due à des poussières interstellaires, car la couleur de la galaxie est très rougie, mais, ce qu'il y a de très curieux, c'est que, malgré la présence de ces poussières, on n'observe aucune étoile très brillante et jeune des types O et B, alors qu'en général, dans notre Galaxie et les autres galaxies normales, à poussière est associé gaz et à gaz est associée étoile jeune ayant pris naissance en son sein; par exemple, dans la nébuleuse d'Andromède, la majorité de la luminosité des bras spiraux provient de ces étoiles récemment formées à partir du gaz. Cette morphologie très spéciale vient d'être complétée par la découverte d'une très forte polarisation optique, atteignant par endroits 15 p. 100.

On pense que cette polarisation est due à un champ magnétique d'aspect dipolaire dont l'axe serait perpendiculaire au plan principal de cette galaxie. Cette morphologie est unique pour l'instant dans les annales des galaxies. Au point de vue radio-électrique, l'aspect le plus remarquable de Messier 82 vient de ce que son spectre radio est le plus plat que l'on connaisse parmi les sources extra-galactiques. Seul lui ressemble le spectre de la nébuleuse du Crabe, reste de supernovae que l'on sait être le siège d'un champ magnétique intense. Il a été possible de préciser la position et la forme de la source Messier 82 : une source circulaire, d'un diamètre de l'ordre de 1′ d'arc, centrée au milieu de l'image optique. Etant donné la distance de Messier 82, cette source radio aurait un diamètre de 1 500 années de lumière, comme celui des régions très spéciales occupant le centre de notre Galaxie, mais son pouvoir émissif est beaucoup plus fort. Enfin, par la raie rouge Hα, on a observé que le centre de Messier 82 était occupé par un immense nuage d'hydrogène ionisé ayant l'aspect d'un reste d'explosion gigantesque. Que se passe-t-il dans ce centre d'où semblent s'échapper de fines aigrettes lumineuses?



La nébuleuse Messier 87 ;
ci-dessous, région centrale.



Virgo A et son jet mystérieux

Messier 87 (ou Virgo A) est une galaxie elliptique géante, 150 000 années de lumière de diamètre, et distante de nous de 50 millions d'années de lumière. Pas de bras, pas de matière absorbante; une énorme agglomération d'étoiles très concentrée vers son centre. La photographie sous-exposée représente les régions centrales de Messier 87 à une échelle beaucoup plus grande. Ici apparaît un objet astronomique absolument exceptionnel : le jet, suite de gouttelettes pratiquement alignées qui semblent être éjectées du noyau de la galaxie; ce jet a 3 000 années de lumière de longueur, mais il n'est pas composé d'étoiles, le spectre de sa lumière est continu, sans aucune raie, et, de plus, sa lumière est polarisée à un très fort taux. La seule explication possible est que cette lumière est d'origine synchrotron. Ici les électrons relativistes ont une énergie si grande, ou le champ magnétique dans lequel ils circulent est si intense, que les longueurs d'onde émises tombent dans le domaine optique. Si ce jet ne partait pas exactement du centre de la galaxie, on serait tenté de croire qu'il est le résultat catastrophique de quelque expérience malencontreuse entreprise par des êtres possédant une technologie « avancée ». Le rayonnement radio très intense de cette source se compose de deux parties : premièrement, d'un halo radio-électrique de dimensions comparables à celles de la galaxie optique et de spectre non thermique; deuxièmement, d'une source très petite située en son centre et dont l'émission sur ondes décimétriques est prépondérante. Des mesures interférométriques à haute résolution ont montré que cette source centrale était double, composée de deux centres d'émission séparés l'un de l'autre d'une demi-minute d'arc, soit environ 4 000 années de lumière. Cette source est évidemment à associer au jet optique, mais quelle relation exacte existe-t-il entre ces deux phénomènes? Le mystère est complet pour l'instant.

Cygnus A, radiogalaxie double

Autre exemple plus précis encore : la radiosource Cygnus A. C'est la plus intense radiosource du ciel et l'une des premières à avoir été identifiée à un objet optique. Cette galaxie est très lointaine, puisque, ici, elle a une magnitude photographique 18. Elle consiste en un halo elliptique allongé dans la direction nord-sud et ayant une longueur d'environ une demi-minute. Au milieu de ce halo se trouve un noyau beaucoup plus intense, coupé en deux par une bande absorbante. Le spectre optique de cette galaxie est spécial; il contient la raie de l'oxygène ionisé 3727 très intense. La galaxie est à environ 500 millions d'années de lumière de nous. La structure radio de cet objet est une des mieux connues qui soit. Elle consiste essentiellement en deux sources situées sur une ligne est-ouest et séparées l'une de l'autre de 1′ 30″ d'arc, chaque source ayant un diamètre de 30″ et des intensités comparables. Ces deux globules émettent plus intensément du côté extérieur et sont reliés l'un à l'autre par une zone d'émission plus faible. On a l'impression que ces deux zones ont été éjectées du centre de la galaxie. On est étonné par la différence de structure constatée entre l'objet optique et l'objet radio. De plus, l'objet radio est très grand — il correspond à une dimension d'environ 200 000 années de lumière —, tandis que la galaxie optique a une dimension de l'ordre de 100 000 années de lumière seulement. Dans la majorité des cas, les radiogalaxies se révèlent être doubles. La source centrale de Virgo A serait-elle le stade initial d'un processus dont Cygnus A serait le stade final? Peut-être. Il existe même certaines sources, tel Centaurus A, qui se composent au point de vue radio de deux sources doubles, un peu comme si l'on superposait l'image de Cygnus A à celle de Virgo A.

Bootes A, la plus lointaine radiosource identifiée

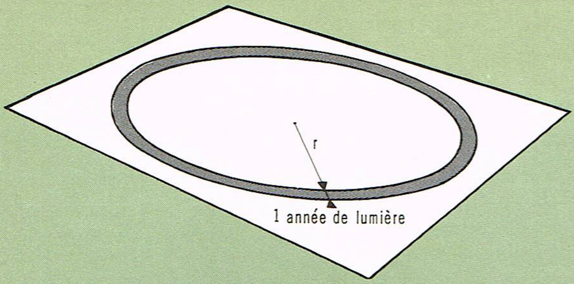
**La radiogalaxie la plus lointaine qui ait été identifiée est Bouvier A.
En effet, bien que l'on connaisse maintenant des milliers de radiosources extra-galactiques, on n'a pu en identifier encore optiquement que quelques dizaines; cela tient au fait que les radiogalaxies sont des objets intrinsèquement très puissants et que la majorité d'entre elles sont à des distances tellement grandes dans les profondeurs de l'univers que les galaxies optiques qui doivent probablement leur correspondre sont pratiquement en dehors de la portée des télescopes optiques actuels. Son intensité radio est encore très respectable, un centième de celle de Cygnus A. Grâce à des mesures de position précises, il a été possible de l'identifier avec une galaxie de magnitude 20. Elle est située dans un amas d'un diamètre de 3′ d'arc contenant une soixantaine de galaxies, à la limite de visibilité du télescope du Mont-Palomar. La galaxie la plus brillante est en réalité située au-devant de cet amas, comme le prouve la mesure de sa vitesse radiale de récession. La deuxième plus brillante est celle qui a été identifiée à la source Bouvier A. Son spectre a révélé le fait surprenant que la raie de l'oxygène ionisé de 3 727 Å, raie ultraviolette, est décalée par l'expansion de l'univers dans la région verte du spectre de 5 500 Å. C'est donc à la vitesse fantastique de près de 150 000 km/s que cette galaxie s'éloigne de nous, soit la moitié de la vitesse de la lumière. Sa distance est estimée à plus de 5 milliards d'années de lumière. On voit donc cette galaxie dans l'état où l'univers était il y a 5 milliards d'années; la Terre n'était pas encore née et le Soleil commençait peut-être seulement à exister. Un interféromètre à très haute définition constitué de deux antennes situées à 50 km l'une de l'autre a montré que cette radiosource était elle aussi composée de deux globules situés à 4″ d'arc seulement l'un de l'autre, ce qui correspond à une dimension d'environ 300 000 années de lumière, comparable à celle de Cygnus A.**

la répartition des radiogalaxies

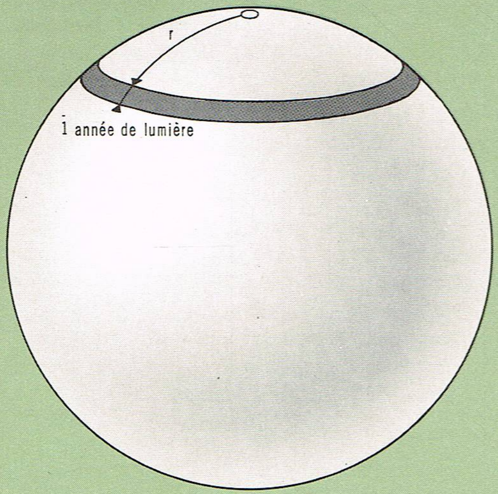
On a pu calculer les puissances radio intrinsèques des sources pour lesquelles une identification a été faite, puisque, dans ce cas, on connaît la distance de l'objet. Il a été ainsi possible de construire ce que l'on appelle la « fonction de luminosité radio » des galaxies. Lorsque l'on connaît la distance d'une radiogalaxie et sa magnitude radio apparente, il est possible de calculer sa magnitude radio absolue, qui, comme en optique, donne une mesure de sa puissance émettrice. La fonction de luminosité radio donne le nombre de galaxies plus puissantes que la magnitude M_r contenue par mégaparsec cube d'espace. Les magnitudes radio absolues s'étendent sur un très large intervalle. Les deux plus puissantes, celles de Cygnus A et Bouvier A, ont la magnitude absolue radio — 35. Les plus faibles se situent aux alentours de — 20, — 22, telle, par exemple, celle de Messier 82. Les plus intenses sont les moins nombreuses. Des radiogalaxies telles que Cygnus A ou Bouvier A sont séparées l'une de l'autre de 3 milliards d'années de lumière en moyenne. Celles qui, en intensité, sont comparables à Virgo A se rencontrent tous les 300 millions d'années de lumière, et les plus faibles, tous les 30 millions d'années de lumière environ. Rappelons que les galaxies normales sont séparées, elles, par quelques millions d'années de lumière. Présenté d'une façon différente, le résultat est qu'environ une galaxie sur un milliard est radiosource du type de Cygnus A et une sur un million du type de Virgo A.

Ces résultats sur la fonction de luminosité radio sont fondamentaux pour l'interprétation des comptages de radiogalaxies en fonction de leur intensité apparente, car ce sont eux qui donneront l'échelle des distances. Pour deux radiogalaxies observées avec le même flux apparent, si l'une a un flux intrinsèque quatre fois plus grand que l'autre, elle sera deux fois plus loin. La forme de la fonction de luminosité est telle que, lorsque l'on regarde jusqu'à une limite de flux donnée, par exemple jusqu'au pouvoir de sensibilité maximal d'un radiotélescope donné, la majorité des sources observées sont en réalité des sources intrinsèquement très puissantes situées à une très grande distance. Une minorité seulement sera constituée par des radiogalaxies intrinsèquement faibles situées près de nous.

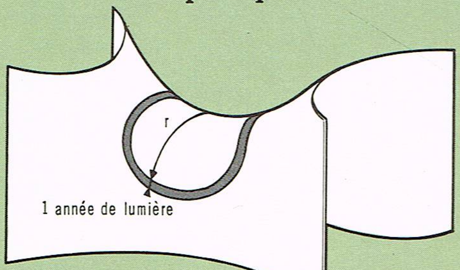
Les comptages de galaxies dans les trois géométries :



euclidienne

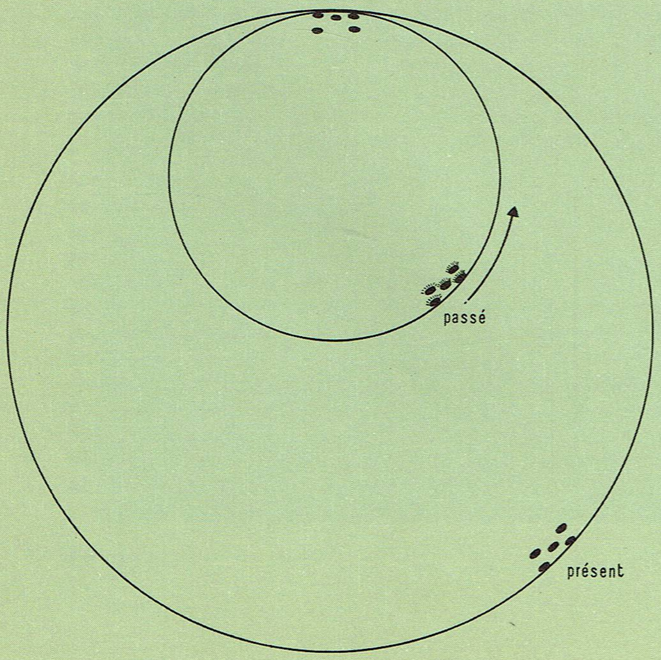


sphérique



hyperbolique

Influence de l'expansion et de l'évolution dans les comptages de galaxies.



l'évolution de l'univers

On saisit le pouvoir de pénétration énorme offert par la radio-astronomie pour sonder l'univers. Mais ici interviennent, dans l'interprétation des résultats, des notions très compliquées d'effets relativistes, d'effets de relativité générale. On les saisit déjà en pensant au fait que Bouvier A s'éloigne de nous avec une vitesse égale à la moitié de celle de la lumière. En regardant seulement un peu plus loin, on doit avoir affaire à des objets dont la vitesse devient comparable à celle de la lumière; il faut donc tenir compte des effets relativistes. **Autre fait évident : de nombreux indices observationnels semblent indiquer qu'il y a environ une dizaine de milliards d'années il s'est passé quelque chose d'important dans l'univers.** Cela ne signifie pas nécessairement qu'il ait été créé à ce moment; il se peut, plutôt, qu'il soit passé par un minimum de contraction avant de reprendre son expansion, ou alors que, tout en ayant un âge infini, il ait un âge moyen de l'ordre de 10 milliards d'années. Notre Galaxie, elle aussi, semble avoir un âge comparable à cette valeur. Or Bouvier A est observé dans l'état où il était il y a 5 milliards d'années; cela fait concevoir que d'énormes effets évolutionnels doivent être introduits dans les données expérimentales pour qu'elles puissent être interprétées correctement, non seulement parce que les galaxies elles-mêmes ont dû évoluer beaucoup, mais aussi parce que l'univers, dans son ensemble, a dû parcourir une longue période de son histoire. **En bref, la cosmologie devient maintenant une science d'observation.**

Au point de vue théorique, le fondement le plus sûr de la cosmologie est la relativité générale d'Einstein, dont il existe plusieurs vérifications expérimentales très convaincantes : l'égalité de la masse pesante et de la masse inerte, l'avance du périhélie de Mercure, la déviation de la lumière au voisinage du Soleil, le déplacement des raies spectrales dans un champ de gravitation, tout particulièrement vérifié très récemment avec une grande précision par l'effet Mossbauer. Les équations de la relativité générale, vérifiées ainsi « localement », pourrait-on dire, ont le grand mérite d'apporter sans aucune hypothèse supplémentaire des modèles d'univers. L'idée fondamentale de la relativité générale est de faire dépendre les propriétés géométriques de l'espace-temps de la distribution de la matière et de l'énergie. Connaissant cette distribution dans l'univers, il serait en principe possible de résoudre les équations, pour obtenir d'une part la géométrie de l'espace, d'autre part son évolution au cours du temps, et pour répondre par exemple aux deux questions fondamentales suivantes : l'espace est-il euclidien, sphérique ou hyperbolique? l'univers est-il en expansion indéfinie ou est-il en oscillation? Pratiquement, pour résoudre ces équations on est obligé d'apporter des hypothèses simplificatrices, par exemple de supposer que l'ensemble de la matière dans l'univers est éparpillé uniformément au lieu d'être distribué sous forme de galaxies (ce n'est pas là une restriction fondamentale). D'autre part, l'observation de l'univers jusqu'à la distance du Bouvier montre que la distribution à grande échelle de la matière est quasiment uniforme. Il n'existe pas, par exemple, une moitié d'univers où il y aurait dix fois plus de galaxies que dans l'autre. On peut donc, dans une première approche, adopter l'hypothèse de l'uniformité. Il en résulte que les équations de la relativité générale se résolvent alors immédiatement et donnent des modèles d'univers dits « uniformes ». Un tel modèle est géométriquement défini par deux grandeurs : premièrement une constante k qui peut prendre les valeurs 0, + 1 ou — 1, selon que la géométrie de l'espace est euclidienne, sphérique ou hyperbolique; deuxièmement une fonction $R(t)$ qui donne l'échelle ou le rayon de courbure de l'espace en fonction du temps. Deux grandeurs physiques viennent compléter la description de l'univers : premièrement sa densité; deuxièmement sa pression. Cette notion de pression ne doit pas sembler particulièrement spéciale; il s'agit de la pression des radiations contenues dans l'univers d'une part et, d'autre part, de la pression due à

l'agitation cinétique des galaxies, qui, au fond, sont les molécules de cet univers. Dans l'état actuel, d'ailleurs, cette pression est négligeable. La relativité générale fournit alors deux équations différentielles entre pression, densité et rayon de courbure; il faut leur adjoindre une équation d'état reliant pression p et densité comme dans toute théorie de gaz. Pour simplifier la description, on peut supposer que cette équation d'état se réduit à $p = 0$. Dans ces équations interviennent des paramètres : la constante de gravitation, qui est connue; la constante k ci-dessus; et la constante cosmologique, qui est inconnue et sera à déterminer expérimentalement. Ce système d'équations peut se résoudre et conduit à un modèle d'univers bien défini si l'on connaît les conditions actuelles, c'est-à-dire densité actuelle, rayon de courbure actuel, valeur de k, et aussi valeur de la constante cosmologique. La densité actuelle de l'univers est une des valeurs les plus importantes à déterminer. Quant au rayon de courbure et au paramètre k, leur détermination directe n'est pas commode, et l'on a coutume, dans le dépouillement des observations expérimentales, de les remplacer par la constante de Hubble H et par le paramètre d'accélération q_0. On sait qu'actuellement l'univers est en expansion; la vitesse de cette expansion est approximativement connue et fournit donc la valeur du paramètre de Hubble. Le paramètre d'accélération q_0 est beaucoup plus difficile à déterminer, car, en fait, il représente la vitesse à laquelle varie l'expansion de l'univers. Pour le déterminer, il faudrait, par exemple, mesurer la constante de Hubble dans la région de Bouvier A, ce qui est techniquement impossible.

Quant à la constante k, c'est-à-dire la nature de la géométrie de l'espace, on peut en principe la déterminer par comptage des radiosources en fonction de la distance. Nous pouvons obtenir une représentation imagée commode si nous nous restreignons à un espace statique à deux dimensions. La géométrie euclidienne est alors celle d'un plan, et si, à partir de notre Galaxie, nous traçons une couronne de rayon r et de largeur une année de lumière, le nombre de sources qui y seront contenues sera proportionnel à r. Si l'univers au contraire est sphérique, la géométrie est celle de la surface d'une sphère, où les droites, les plus courts chemins d'un point à un autre, sont représentées par des arcs de grand cercle. Là le nombre de sources croîtra moins vite que r. Au contraire, dans la géométrie hyperbolique représentée par la surface selle de cheval, ce nombre croîtra plus vite que r. La réalité est cependant plus compliquée, car l'univers n'est pas statique, et des effets parasites, calculables d'ailleurs, vont venir se superposer. Par exemple, dans un modèle d'univers sphérique en expansion, deux effets vont intervenir : au voisinage de notre Galaxie, en O, nous voyons les sources dans leur état de brillance et de dispersion actuel, mais, loin de nous, nous les voyons dans l'état passé où, premièrement, elles étaient plus concentrées, et où, deuxièmement, leur émission intrinsèque était peut-être différente. Que le lecteur nous excuse de ne pas pousser plus loin l'analyse de ces situations où la notion de distance même perd sa signification habituelle.

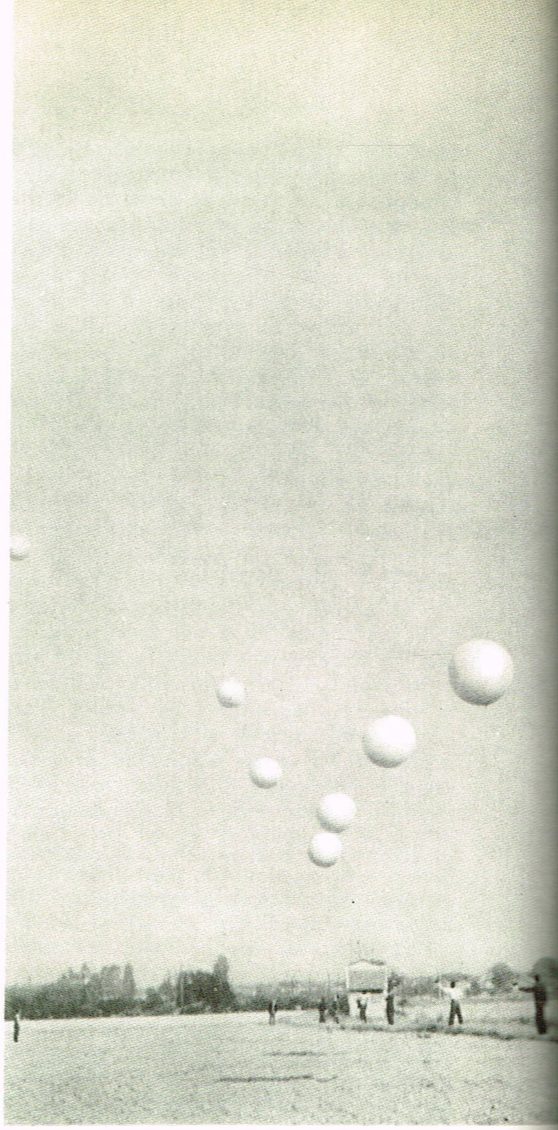
Lorsque des effets relativistes aussi importants interviennent, la distance purement spatiale n'a pas de sens, car le seul véritable invariant est la distance à quatre dimensions dans l'espace-temps, comme on peut déjà s'en rendre compte en étudiant la relativité restreinte. Telle est la raison, d'ailleurs, pour laquelle il est impossible d'observer une galaxie qui s'éloigne de nous à une vitesse plus grande que la vitesse de la lumière.

En conclusion, la radio-astronomie a le pouvoir de pénétration suffisant pour décider de la structure de l'univers. Mais, auparavant, il lui faut connaître la nature intime des radiogalaxies afin d'être capable de tenir compte dans ses sondages des effets évolutifs.

Nous jonglons en ce moment avec des distances de l'ordre du milliard d'années de lumière. Essayons de revenir en arrière par bonds successifs. Réduisons 10 000 fois l'échelle de nos préoccupations; nous arrivons à 100 000 années de lumière, les dimensions de notre Galaxie. Réduisons 10 000 fois encore, nous arrivons aux étoiles proches; 10 000 fois encore, nous voici dans le système solaire; 10 000 fois encore, nous voici à la distance de la Lune. Arrêtons-nous là, car, aussi bien, la Lune, c'est presque chez nous.

LES RAYONS COSMIQUES

par
Louis LEPRINCE-RINGUET



Le premier messager qui nous ait fourni une quantité considérable d'informations sur le Soleil, les planètes, les galaxies, les nébuleuses de toute sorte, c'est le rayon lumineux : en analysant la lumière émise par une étoile, on peut déterminer sa température, sa composition, la vitesse relative de l'étoile par rapport à la Terre, et bien d'autres caractères de l'objet céleste, comme nous l'a montré le premier chapitre.

La lumière est à coup sûr le messager le plus extraordinaire que nous ayons à notre disposition. Mais nous savons également qu'il existe d'autres informations, transmises par des rayonnements électromagnétiques de beaucoup plus grandes longueurs d'onde; le développement merveilleux de la radio-astronomie réside précisément dans la découverte progressive des informations transportées par des ondes radio qui sont émises par des étoiles ou même par des régions de l'espace où nulle étoile n'est visible. Ainsi, la fameuse raie de 21 cm, dont il a été abondamment question dans le chapitre précédent, nous apporte des informations sur les masses d'hydrogène lointaines, même lorsqu'elles nous sont cachées par d'autres masses d'hydrogène plus proches, en mouvement relatif par rapport aux premières.

Mais il est encore un troisième messager dont on a beaucoup parlé depuis 1912, date de sa découverte : c'est le rayon cosmique. Il ne s'agit plus d'onde électromagnétique, comme la lumière et l'onde radio; il s'agit de particules, et même de particules nucléaires animées de vitesses extrêmement grandes. Si la plupart d'entre elles sont constituées par des protons, on y trouve aussi des noyaux d'hélium, de carbone, d'oxygène, d'azote et d'autres éléments jusqu'au fer. En gros, on peut dire que la répartition des particules primaires du rayonnement cosmique correspond à peu près à la répartition des éléments dans les étoiles.

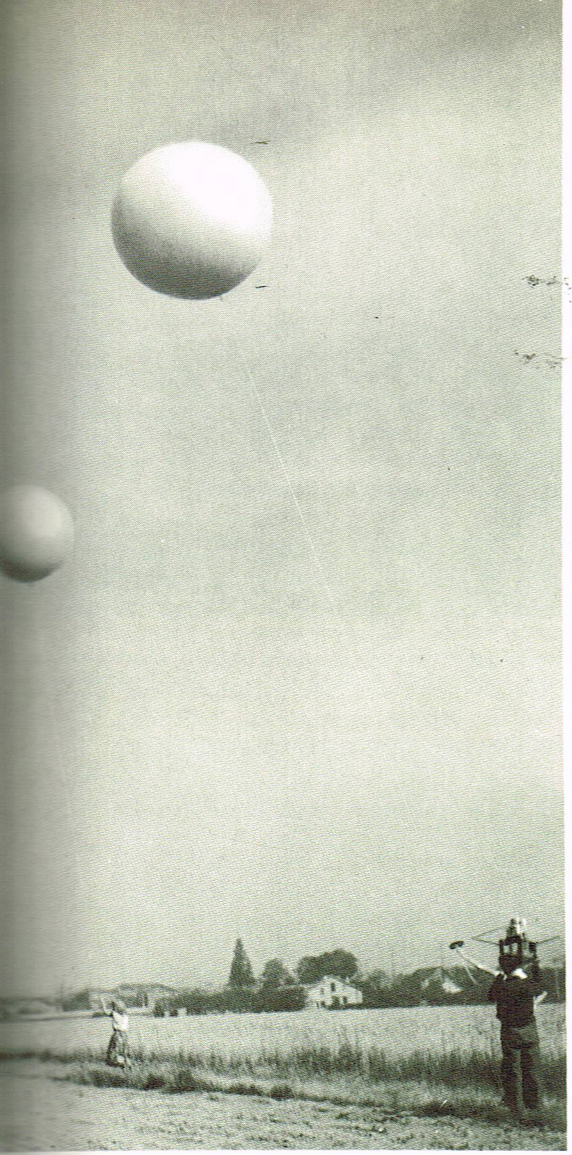
Nous allons voir comment l'on a fait progressivement connaissance de ces rayonnements et ce que l'on peut en extraire comme informations.

découverte du rayonnement cosmique

C'est en 1912 que le physicien autrichien Viktor Hess a mis en évidence un rayonnement très pénétrant, dont l'intensité augmente quand on s'élève dans l'atmosphère.

L'histoire de cette découverte est bien intéressante, d'ailleurs, d'un point de vue général. En effet, dans les chambres d'ionisation les mieux isolées, au niveau du sol, on décelait toujours un très petit courant qui déchargeait les électromètres. On pouvait penser que ce petit courant était dû à un rayonnement provenant de la radioactivité de la Terre, par exemple, et on voulait s'en débarrasser. C'est pour cela que Hess fit monter ses appareils en ballons, et le résultat fut surprenant puisque le rayonnement augmentait quand on s'éloignait de la Terre; il y avait donc autre chose que ce qui était prévu, et, progressivement, à la suite de nombreuses expériences faites en Autriche, et effectuées

Phot. André Steiner.



Lâcher de ballons par l'équipe de l'École polytechnique. Chaque ballon emporte un paquet de plaques photographiques dont l'émulsion gardera l'image latente des désintégrations provoquées par l'impact des particules cosmiques de haute énergie.

également en Amérique par Millikan, on a découvert qu'il s'agissait d'un rayonnement très pénétrant venant de l'extérieur de l'atmosphère. On l'appela « cosmique » pour bien indiquer que ce rayonnement provenait de très loin, sans que l'on pût dire naturellement quelle en était l'origine, problème qui n'est d'ailleurs pas résolu actuellement dans toute sa complexité.

On ne pouvait pas, à l'époque, détecter chacune des particules constituant ce rayonnement; c'était un effet global qu'on pouvait analyser, et les progrès furent très lents jusque vers 1931, époque à laquelle les physiciens réussirent à mettre au point des détecteurs de particules individuelles. On verra plus loin (tome II) que la méthode de la chambre de Wilson, celle des compteurs Geiger, celle des émulsions photographiques et également celle des amplificateurs ultra-sensibles permettent de détecter les rayonnements individuels. Le passage d'une particule très rapide dans un appareil produit de l'ionisation, que l'on détecte, et l'on peut, par exemple au moyen de compteurs d'électrons en coïncidence, enregistrer le passage de toutes les particules traversant l'ensemble de ces compteurs.

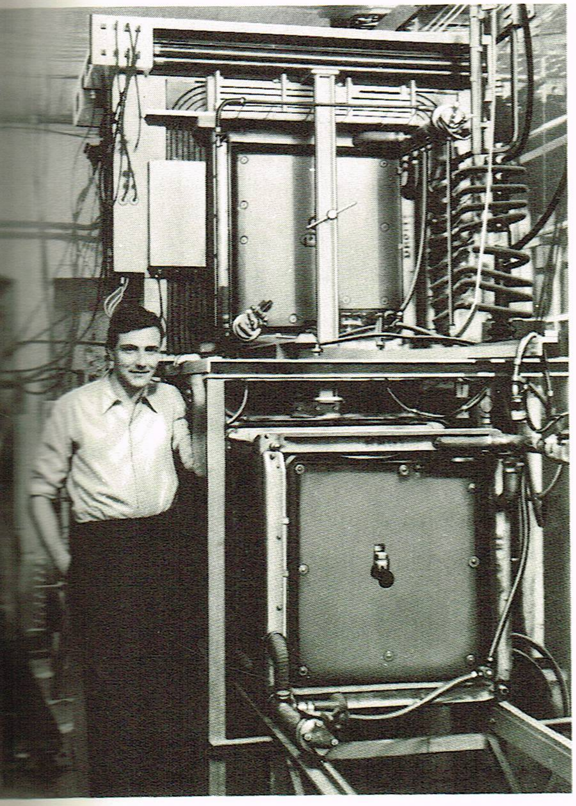
Dans l'histoire du rayonnement cosmique, la période la plus importante fut certainement celle qui débuta vers 1930 et se termina vers 1960. Pendant ces trente années, l'on réussit à avoir des informations de plus en plus précises sur la composition des rayons cosmiques, sur leur nature, sur leur énergie et, également, sur les effets qu'ils sont susceptibles de produire lorsqu'ils rencontrent des noyaux d'atomes.

Il n'était pas évident que les particules enregistrées correspondaient aux rayons cosmiques primaires : on enregistre bien les particules ionisantes, mais elles peuvent être produites par autre chose, par exemple par des rayons γ de très grande énergie, et l'on a été très longtemps sans savoir si le rayonnement cosmique était électromagnétique ou corpusculaire. Ce problème a été résolu vers 1934 par le physicien hollandais Clay, puis par d'autres équipes de physiciens français et américains. Si le rayonnement cosmique est constitué de photons, c'est-à-dire s'il est de nature électromagnétique, l'action du champ magnétique terrestre doit être nulle, c'est-à-dire qu'on doit observer autant de rayonnement à l'équateur et sous les latitudes élevées. Bien des calculs ont été effectués, notamment par Störmer, puis par Lemaître et Vallarta, sur les trajectoires de particules électrisées très rapides qui se dirigent vers la Terre : au voisinage du pôle, le champ magnétique terrestre n'agit pas, et son action est la plus grande dans le plan équatorial. Autrement dit : des particules venant vers la Terre ont plus de difficulté à atteindre l'équateur que le pôle, si bien que seules des particules dépassant une certaine énergie, que l'on peut calculer, sont capables d'atteindre l'équateur.

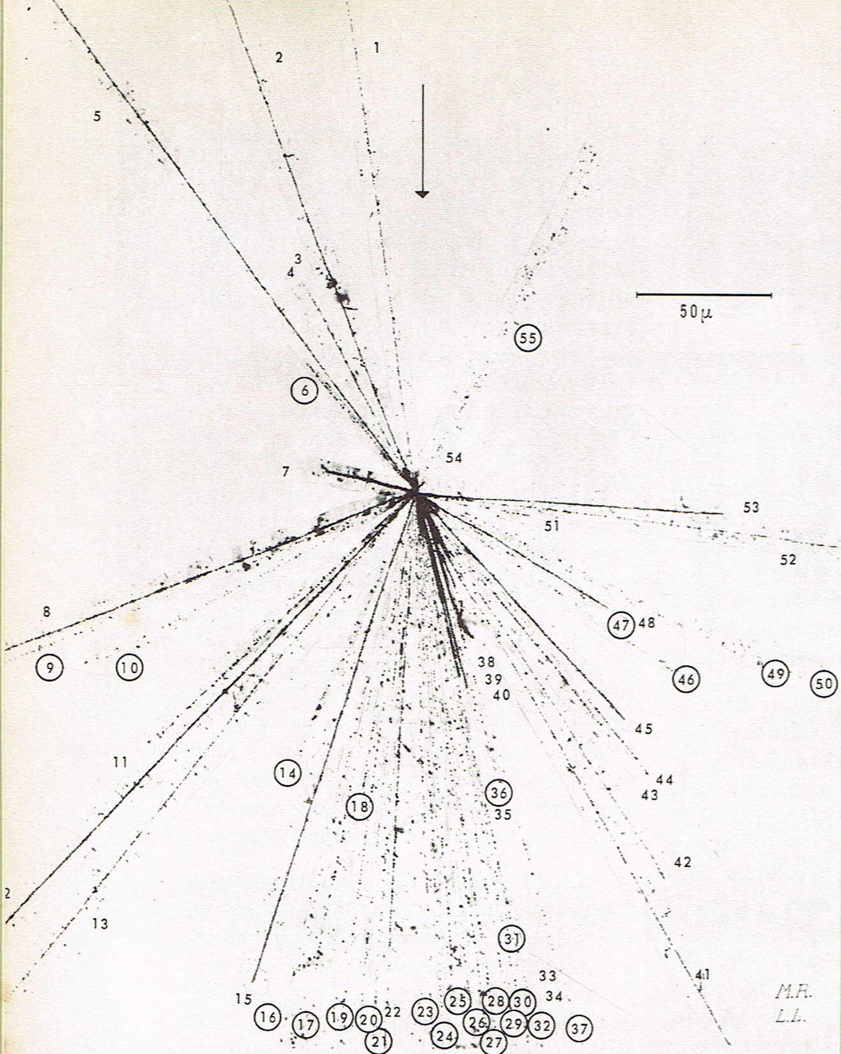
Ainsi, pour savoir si le rayonnement cosmique primaire est électromagnétique ou corpusculaire, il faut faire des expériences en latitude; le mieux serait de les effectuer à la limite supérieure de l'atmosphère, pour éviter les complications de l'absorption du rayonnement cosmique lorsqu'il traverse l'atmosphère, mais on peut également avoir des informations au niveau de la mer. C'est entre 1934 et 1938 que ces informations ont été obtenues à basse altitude, puis à altitude de plus en plus élevée; on a effectivement trouvé une diminution du rayonnement dans les régions équatoriales, ce qui a prouvé sa nature corpusculaire.

le rayonnement cosmique traverse l'atmosphère

Ces particules primaires viennent de toutes les directions, mais, au niveau du sol, elles arrivent filtrées par l'atmosphère, et nous en observons les effets secondaires ou tertiaires lorsque nous cherchons à déceler les rayonnements cosmiques à basse altitude. Une des premières constatations qu'on peut faire facilement, quand on place par exemple le système de compteurs d'électrons au niveau du sol, est que les particules sont en moyenne très pénétrantes : la moitié d'entre elles est capable de traverser un mètre de plomb. On constate aussi que ces particules ne sont pas très nombreuses : le flux est extrêmement faible; une surface égale à la paume



Une des grandes chambres de Wilson installées depuis 1950 au pic du Midi de Bigorre par le groupe de l'École polytechnique.



Grande étoile nucléaire représentant la désintégration d'un noyau d'argent de l'émulsion photographique par une particule neutre d'énergie considérable. Les traces les plus noires sont celles des protons du noyau; le faisceau de traces fines se dirigeant vers le bas est constitué par des mésons. De nombreuses particules neutres (neutrons, mésons neutres, rayons gamma), dont les traces ne sont pas visibles, sont également émises au cours de cette intéraction nucléaire de haute énergie. *(Doc. École polytechnique.)*

de la main en reçoit une par seconde, ce qui est très peu. Cette arrivée est en général très régulière, sauf lors de quelque importante perturbation solaire : jour et nuit c'est la même cadence, qui d'ailleurs ne varie pas de plus du millième selon la saison, la position des étoiles, le cycle solaire, etc. *Nous sommes donc baignés dans un flux de particules primaires, dont nous observons les sous-produits dans l'atmosphère.* Lorsque l'on s'élève en altitude, le nombre des particules reçues augmente beaucoup; il y en a plus de cent fois la dose au sol lorsque l'on atteint 20 ou 30 km d'altitude. Lorsqu'on monte encore, leur nombre diminue un peu, et, finalement, hors de l'atmosphère on arrive aux rayons primaires eux-mêmes.

On a progressivement réussi à connaître l'énergie des rayons cosmiques : elle se compte en milliards d'électron-volts, alors que les particules α (v. chap. *Le monde agité des particules élémentaires*, tome II) du radium ou des corps radio-actifs ont des énergies de quelques millions d'électron-volts seulement. Ainsi l'on a disposé d'une source naturelle de particules d'une énergie très supérieure à celles dont on pouvait disposer dans un laboratoire. *Cela s'est avéré extrêmement intéressant, car on a pu étudier les effets nucléaires des particules de très grande énergie grâce au rayonnement cosmique, bien avant que les grands accélérateurs de particules aient été construits.*

Ce n'est que vers 1960 que les deux ou trois principaux synchrotrons sont entrés en service; comme les accélérateurs fournissent des faisceaux très intenses, des milliards de fois plus intenses que ce que l'on peut obtenir avec le rayonnement cosmique, l'intérêt des physiciens s'est tourné alors vers l'utilisation des grands accélérateurs. Les « cosmiciens » sont descendus de leurs montagnes, ont abandonné leurs expériences en ballons-sondes pour aller s'enfermer dans les grands halls qui entourent les synchrotrons.

Tout cela n'est pas rigoureusement exact. En effet, les particules du rayonnement cosmique ont des énergies très diverses; pour la plupart il s'agit de quelques milliards d'électron-volts, mais certaines ont des énergies beaucoup plus élevées : plus on monte en énergie, plus le nombre de rayons cosmiques est faible, naturellement.

Néanmoins, l'on observe encore dans quelques grands centres des rayonnements cosmiques d'énergie extrêmement élevée, et l'on cherche même actuellement à connaître le maximum de cette énergie. On sait que certaines particules dépassent très largement les mille milliards d'électron-volts; **on a même réussi à observer quelques gerbes produites par des particules dont l'énergie atteint ou dépasse 10^{18} électron-volts, c'est-à-dire un milliard de milliards d'électron-volts.**

Ainsi les rayons cosmiques ont encore un intérêt en tant que particules rapides, et l'on peut essayer d'étudier leurs effets sur la matière. Le rayonnement cosmique est le seul qui puisse nous fournir, très rarement d'ailleurs, des particules dont l'énergie dépasse énormément celle des accélérateurs.

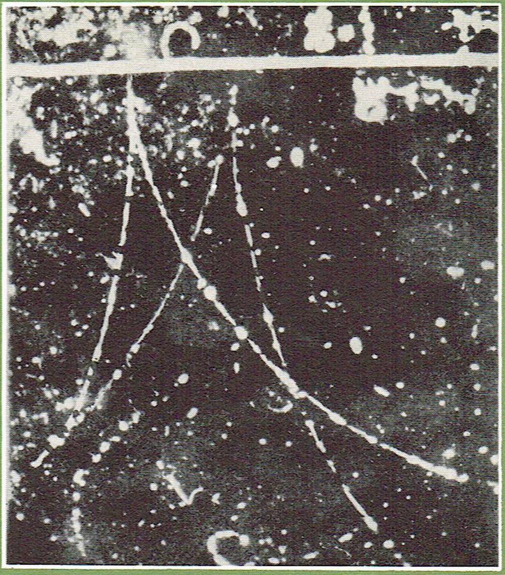
Revenons au passage des rayons cosmiques dans l'atmosphère. Ce que nous observons au niveau du sol, ce sont essentiellement des mésons μ avec quelques électrons rapides. Ces mésons μ sont d'ailleurs capables de s'enfoncer dans le sol et d'être encore observables sous plusieurs centaines de mètres de croûte terrestre. Quand on s'élève dans l'atmosphère, le rayonnement n'a plus la même composition : la proportion des mésons μ diminue et celle des électrons augmente. D'autre part, on observe alors des nucléons rapides. Il y a en effet de plus en plus de protons et de neutrons de grande énergie quand on s'élève au-dessus du sol.

Pour étudier l'interaction du rayonnement et des noyaux, les physiciens ont installé de grands laboratoires dans les observatoires de montagne entre 3 000 et 4 000 m d'altitude. Il y en a aux Etats-Unis, en particulier sur le mont Evans, près de Denver. Il y en a également en Europe dans les Alpes; les laboratoires de la Jungfrau en Suisse, de la Testa Grijia en Italie, de l'aiguille du Midi en France ont fait de remarquables travaux sur l'interaction du rayonnement cosmique et de la matière. Plus récemment, une installation très puissante a fonctionné au pic du Midi de Bigorre avec deux grandes chambres de Wilson superposées, déclenchées par des quantités de compteurs d'électrons. A titre d'exemple, l'installation du pic du Midi a fonctionné jour et nuit sans interruption pendant plus de six ans.

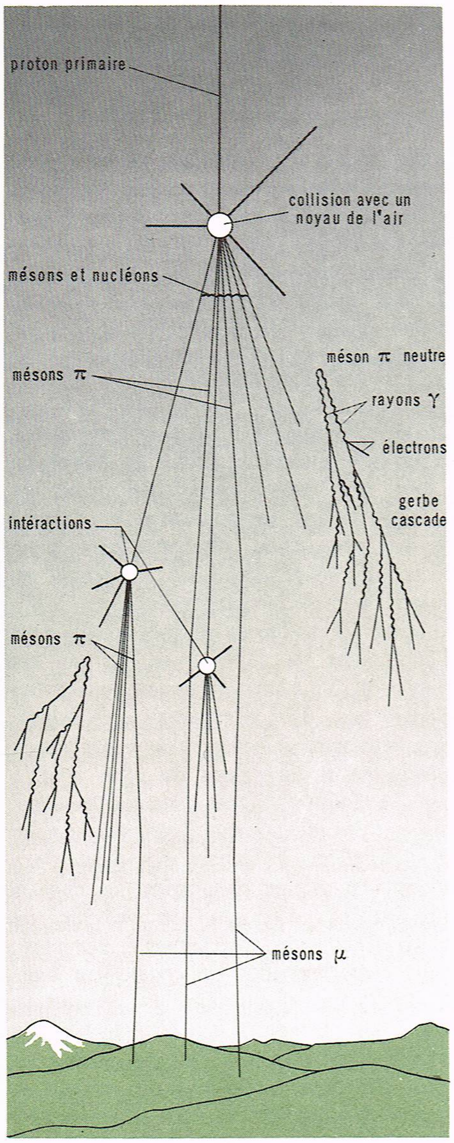
On peut se représenter maintenant assez bien le comportement du rayonnement cosmique à travers l'atmosphère.



Laboratoire du Centre national de la recherche scientifique érigé en 1946-1948 dans le massif du Mont-Blanc, à l'aiguille du Midi (3 650 m), pour l'étude des rayons cosmiques.



Paires d'électrons produites par des rayons gamma de grande énergie provenant du rayonnement cosmique.



Succession d'interactions dues à un proton primaire pénétrant dans l'atmosphère.

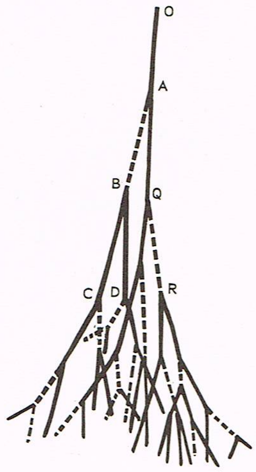
La figure ci-contre montre les effets successifs d'un proton primaire frappant les noyaux de l'atmosphère. Dans la partie très haute, le proton a une grande probabilité de heurter un noyau, par exemple un noyau d'oxygène ou d'azote; comme il possède une très grande énergie, il y aura création possible de mésons π, et peut-être aussi de mésons lourds. D'autre part, le noyau sera complètement disloqué, et les nucléons qui le constituent vont être projetés avec de grandes vitesses vers le bas. Ainsi, après cette première collision, on verra des protons, des neutrons et une gerbe de mésons π s'échapper dans le prolongement de la direction du proton incident. Les protons et les neutrons seront capables, à leur tour, de produire un peu plus bas d'autres interactions moins violentes.

Le comportement des mésons π rapides est très intéressant : en effet, les mésons π créés peuvent être chargés (positifs ou négatifs) ou neutres; ceux qui sont chargés vont se désintégrer au bout d'un temps moyen d'un cent-millionième de seconde. Comme ils vont très vite, ils auront tout de même franchi quelques centaines de mètres, d'autant plus que leur parcours apparent est accru du fait de la relativité.

Les produits de désintégration du méson π vont être des mésons μ; or ces derniers ont une vie moyenne cent fois plus grande : ils sont donc capables de franchir des kilomètres d'atmosphère avant de se désintégrer spontanément. C'est pour cela que les mésons μ arrivent jusqu'au niveau du sol et que nous les observons facilement. D'ailleurs, ils ont une autre particularité : il n'y a pas d'interaction entre eux et les noyaux. *Alors que les mésons π, les protons, les neutrons sont des particules « nucléaires » et interagissent facilement avec des noyaux ou des nucléons, le méson μ traverse la matière sans interaction nucléaire, produisant seulement une minime ionisation le long de son parcours.*

Si la plupart des mésons π se transforment spontanément en mésons μ, certains cependant donnent à leur tour des interactions nucléaires analogues à celles du proton initial, mais, comme ils ont moins d'énergie, ces interactions sont moins importantes, leur nombre dépendant naturellement de la densité de la matière traversée. Si bien que l'on peut dire en gros que les mésons π produits très haut, là où l'atmosphère est très peu dense, se désintégreront presque tous en donnant des μ, alors que les mésons π produits très bas, là où l'atmosphère est très dense, donneront surtout des effets nucléaires.

Nous n'avons pas encore parlé du méson π neutre, dont le comportement est tout à fait particulier. Ce méson se désintègre extrêmement vite, en un temps moyen de 10^{-15} seconde, c'est-à-dire en un millionième de milliardième de seconde. Il n'a donc pas le temps de franchir une longueur appréciable et se transforme en deux rayons γ, qui vont naturellement se diriger vers le bas. Ces rayons γ vont se transformer en paires d'électrons par création de couples électron-positon, et ces électrons vont eux-mêmes, par freinage dans la matière, donner d'autres rayons γ, d'énergie moindre naturellement, qui à leur tour produiront des paires, etc. Si bien que le méson π^0 est à l'origine de ces magnifiques phénomènes que l'on appelle des « gerbes cascades » et qui peuvent être au niveau du sol d'une très grande extension (plusieurs centaines de mètres).



Explication du phénomène des gerbes (en traits pleins, les électrons; en pointillé, les photons) : OA, électron; AB, photon de freinage qui matérialise deux électrons BC et BD; Q, photon de freinage qui perd son énergie en R par matérialisation d'une paire d'électrons. Ces phénomènes se répètent à partir de CD et R, et la gerbe atteint rapidement une grande complexité.

Départ d'un ballon-sonde porteur d'une chambre d'ionisation pour l'étude des rayons cosmiques (université de Princeton).



Tout ce que nous venons de dire est schématisé dans la figure ci-contre. On se rend compte à quel point la compréhension de tous ces phénomènes a intéressé les « cosmiciens », les poussant à inventer de nouvelles méthodes de détection susceptibles d'explorer l'atmosphère à tous ses niveaux.

Il faut encore ajouter que, rarement il est vrai, d'autres particules peuvent également être créées, par exemple les mésons lourds et les hypérons; c'est grâce au rayonnement cosmique d'ailleurs que les mésons lourds et les hypérons ont été découverts depuis la fin de la guerre et que leurs propriétés ont été élucidées. Jusque vers 1960, il s'agissait de phénomènes très rares; actuellement, ce sont les grands accélérateurs qui fournissent les faisceaux de mésons lourds, d'hypérons et même d'antiprotons.

ballons-sondes et rayons cosmiques

Le désir d'observer des rayons cosmiques primaires et d'étudier la première interaction qu'ils effectuent a poussé les physiciens à essayer d'expérimenter à très haute altitude, aux limites mêmes de l'atmosphère. Pour cela, ils ont utilisé essentiellement la méthode des émulsions photographiques épaisses et concentrées. Dans de telles émulsions, les rayons ionisants laissent une trace que l'on peut développer lorsque l'on a récupéré le paquet. *Il s'agit donc d'envoyer un paquet d'émulsions vierges à 30 ou 40 km d'altitude, de le faire plafonner plusieurs heures, de le récupérer après irradiation par le rayonnement primaire.*

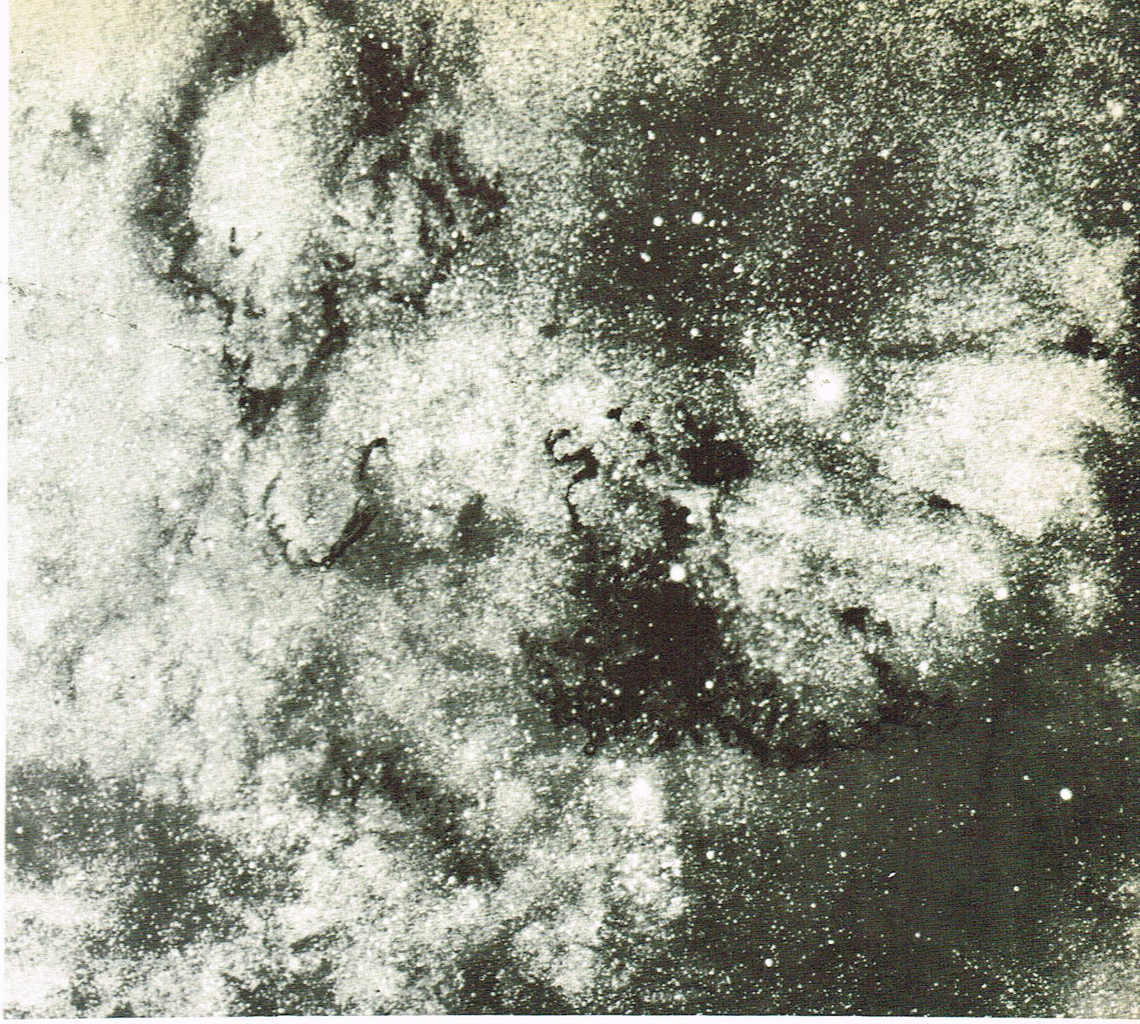
On a commencé par utiliser des ballons-sondes, mais, pour soulever un poids assez considérable, de l'ordre de plusieurs kilogrammes ou même de plusieurs dizaines de kilogrammes, il fallait organiser des chapelets de ballons : technique très intéressante, qui s'est largement développée chez les « cosmiciens » et a permis de faire de nombreuses expositions. Mais on a également réussi à fabriquer de grands ballons extrêmement légers, qui n'étaient plus constitués par du caoutchouc, mais par un plastique, le polyéthylène; ils ont été fabriqués essentiellement aux Etats-Unis, à Minneapolis, et mesuraient jusqu'à 30 m de hauteur. Peu gonflés au départ, ils se dilataient sous l'effet de l'ascension et de la dépression correspondante, plafonnaient pendant longtemps et, grâce à un déclenchement, lâchaient leur charge, qui retombait, freinée par un parachute. On disposait naturellement des émetteurs radio qui permettaient de détecter et de retrouver l'objet à terre.

aspect astrophysique du rayonnement cosmique

Dans tout ce qui vient d'être dit, on considère le rayon cosmique comme une particule qui nous est fournie gracieusement et que nous utilisons dans un but de connaissance nucléaire. Nous avons fermé les yeux jusqu'à présent sur l'origine et le mode de production du rayonnement cosmique, et nous nous sommes préoccupés essentiellement de savoir ce qui se passait lorsqu'un rayon de grande énergie frappait un noyau ou un nucléon.

Mais l'aspect astrophysique de ce rayonnement est du plus haut intérêt. *Comment les rayons cosmiques sont-ils produits, par quel mécanisme sont-ils accélérés, combien de temps passent-ils en moyenne dans la Galaxie avant d'être rencontrés par la Terre?* Autant de problèmes de grande importance, qui ne sont d'ailleurs pas résolus.

On peut évoquer ici un problème particulier très intéressant, celui du temps de séjour des rayons cosmiques entre le moment de leur production et le moment de leur absorption. On peut parvenir à avoir une information sur ce temps en faisant le raisonnement suivant : supposons que les rayons cosmiques proviennent d'explosions de supernovae, ce qui est bien probable pour une bonne proportion d'entre eux. S'ils parcourent une galaxie complètement vide (en étant simplement maintenus à l'intérieur de cette galaxie par des champs magnétiques convenables, susceptibles de courber leurs trajectoires et d'empêcher le départ des particules), on aura la même proportion de noyaux d'hydrogène, d'hélium, de carbone, de fer, que celle qui correspond à leur création; autrement dit, on retrouvera dans le rayonnement cosmique la proportion exacte des éléments des étoiles. Mais la galaxie n'est pas vide; il y a essentiellement de l'hydrogène, et l'on connaît la densité de cette matière. Que se passera-t-il alors? Les noyaux du rayonnement cosmique vont de temps en temps heurter un noyau d'hydrogène de la galaxie; il y aura donc une cassure de ce noyau au



Y a-t-il des rayons cosmiques d'origine extra-galactique ? L'étude des rares phénomènes de très grande énergie permettra probablement de le savoir prochainement, les particules d'énergie égale à 10^{19} eV ne devant pas pouvoir rester longtemps dans la seule Galaxie, dont nous voyons ici un aspect.

profit de noyaux plus petits, si bien que l'on n'observera plus tout à fait la même proportion des éléments les plus lourds. Or, ceci est très intéressant, car, dans les étoiles, il n'y a pratiquement pas de lithium, ni de béryllium, ni de bore, qui sont les éléments de numéro atomique 3, 4, 5. Plus les noyaux du rayonnement cosmique rencontreront d'hydrogène, plus il y aura alors de noyaux de lithium, de bore et de béryllium fabriqués. Par conséquent, on conçoit bien qu'une étude approfondie de la proportion de ces trois éléments dans le rayonnement cosmique puisse donner des indications sur le nombre de noyaux d'hydrogène rencontrés, c'est-à-dire sur le temps de séjour des rayons cosmiques dans la Galaxie. Beaucoup d'expériences ont déjà été faites sur ce sujet; elles ne sont pas encore tout à fait définitives, mais elles semblent indiquer que *le rayonnement cosmique a séjourné dans l'espace interstellaire une quarantaine de millions d'années en moyenne avant de frapper la Terre*.

Un autre problème très important est de savoir si les rayons cosmiques sont d'origine solaire, d'origine galactique, ou s'il y en a de provenance extra-galactique. On sait qu'il y a des rayons cosmiques d'origine solaire; ce sont des rayons d'énergie faible ou moyenne, qui peuvent être produits en particulier lors de certaines perturbations solaires. Un grand nombre de centres de détection existent dans le monde, et lorsqu'il y a une explosion solaire, on voit des variations de rayonnement cosmique assez complexes d'ailleurs liées à cette émission. Mais on sait parfaitement que l'on ne peut pas expliquer ainsi la partie dure des rayonnements cosmiques. En revanche, on peut essayer de savoir s'il y a des rayonnements cosmiques extra-galactiques. Il faut pour cela connaître les plus grandes énergies qui sont atteintes par les particules individuelles de ce rayonnement. Plus l'énergie est grande, plus le rayon de courbure de la trajectoire de la particule est grand, et, à l'aide des champs magnétiques que l'on peut envisager dans la galaxie, on limite ce rayon de courbure à une fraction des dimensions de la galaxie. Si donc l'énergie est supérieure à celle qui correspond à ce rayon de courbure, cela veut dire que la particule est très probablement d'origine extra-galactique. On n'a pas actuellement de réponse décisive, mais il semble bien que le rayon de courbure galactique soit déjà dépassé par quelques phénomènes de très grande énergie.

Quant au mécanisme d'accélération du rayonnement cosmique, il est encore très mystérieux. Certaines théories thermodynamiques ont été imaginées, en particulier par Fermi, pour essayer d'expliquer comment des particules, éjectées à moyenne vitesse lors d'une explosion, peuvent être ensuite accélérées et atteindre des énergies considérables grâce à des interactions entre ces particules et les champs électriques ou magnétiques susceptibles d'exister, grâce également aux interactions entre ces particules et des masses de gaz. Actuellement, on n'a encore aucune assurance à ce sujet; il est probable que les années à venir nous donneront d'importantes informations sur ces mécanismes, à mesure que se développeront nos connaissances de la structure de la Galaxie, des intensités et des étendues des champs électriques et magnétiques que l'on peut y rencontrer.

Phot. E. E. Barnard, observatoire Yerkes.

premiers pas DANS L'ESPACE fusées et satellites

par

Maurice VERNET-LOZET

et

Gabriel de CLAVIÈRE

On a continué de dire que l'événement qui a dominé le XIXe siècle a été, sinon l'invention, du moins le développement de la machine à vapeur. Bien qu'il soit souvent imprudent de procéder à des anticipations de ce genre, on peut, en 1969, penser que, pour les générations futures, l'événement capital de notre siècle aura été les premiers pas effectués par l'Homme en dehors du milieu terrestre qu'il n'avait pas quitté depuis sa création.

La surprise provoquée par le lancement du premier Spoutnik, en octobre 1957, fut trop totale, et le caractère de l'événement trop sensationnel pour que celui-ci soit immédiatement situé dans son contexte historique et technique. On préféra exploiter le côté merveilleux de l'affaire, et l'idée que l'on se fit de l'astronautique à cette époque était sans aucun doute plus proche des aventures de Tintin et du professeur Tournesol dans *Objectif Lune* que de la réalité.

Le temps et le prodigieux développement qu'a connu à cette époque l'astronautique ont opéré leur œuvre de démystification, et l'événement restitué dans son contexte apparaît aujourd'hui comme une de ces manifestations qui concrétisent et jalonnent le progrès continu des techniques.

Sans compter les douze méthodes proposées par Cyrano pour aller sur la Lune, il y a longtemps que les hommes avaient rêvé de s'évader de leur condition de terrestres et, dès le début de ce siècle, Tsiolkovski, Esnault-Pelterie et d'autres avaient indiqué, sommairement, mais avec peu d'erreurs de principe, les moyens d'y parvenir. Si l'astronautique n'est pas sortie du domaine du roman d'aventures et de la science-fiction à cette époque, c'est essentiellement parce que les possibilités offertes alors par les diverses techniques mises en jeu ne le permettaient pas.

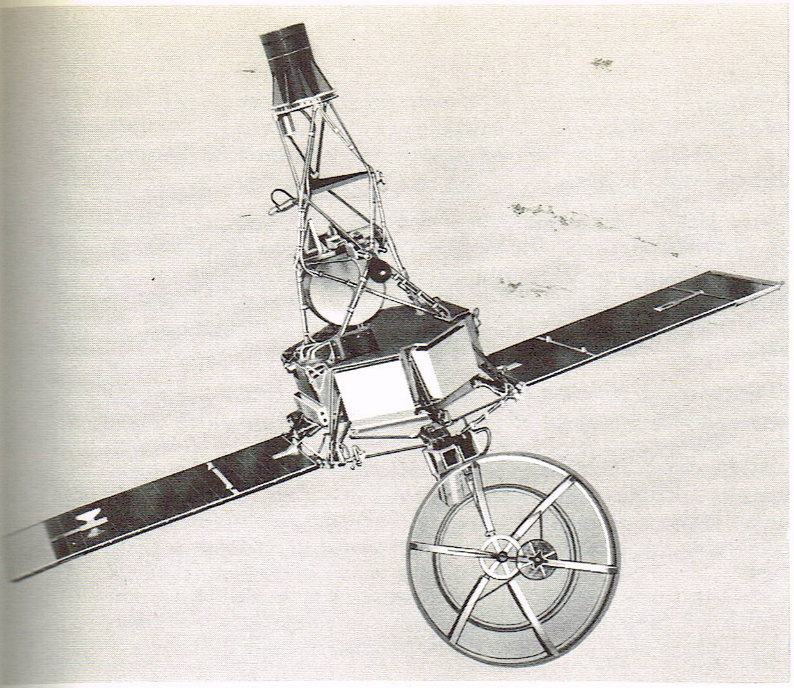
L'astronautique n'est en elle-même ni une science ni une technique, mais elle se situe à l'un des carrefours de nombreuses sciences et techniques, pour ne pas dire pratiquement de toutes.

Pourtant, pour se libérer de l'attraction terrestre, il fallait disposer d'un moyen de propulsion capable :
— de fonctionner sans liaison mécanique avec la Terre;
— de fonctionner en dehors de l'atmosphère (donc sans utiliser l'oxygène de celle-ci);
— de s'accélérer jusqu'à des vitesses 10 à 20 fois plus élevées que celles qui sont atteintes par les avions les plus rapides.

Ce moyen de propulsion est la fusée. Le principe de celle-ci est connu depuis fort longtemps, puisque les Chinois en utilisaient déjà pour augmenter la portée de leurs flèches voici bientôt mille ans.

Mais on a su construire des fusées suffisamment puissantes seulement lorsque les progrès des techniques ont permis de maîtriser l'usage de carburants et d'oxydants suffisamment énergétiques, d'alliages suffisamment légers, de moyens de mesure, de télémesure et de télécommande suffisamment évolués et précis, etc.

Ne pouvant aborder en quelques pages tous les aspects de l'astronautique, ni évoquer tous les problèmes, qui, d'ailleurs, le plus souvent, sont de caractère purement technologique, nous nous limiterons donc à traiter, le plus souvent de façon surtout descriptive, deux sujets. En premier lieu, nous parlerons des fusées, puisque, en fait, c'est la taille et la puissance de celles-ci qui conditionnent l'ordre de grandeur des masses qu'il est possible d'envoyer dans l'espace. En second lieu, nous traiterons des satellites et des missions susceptibles d'être réalisées à l'aide des moyens techniques actuellement à notre disposition.



La sonde spatiale « Mariner ».

fusées

une fusée, pourquoi ?

La mécanique nous apprend (v. tome II, chap. *La mécanique à l'heure spatiale*) que pour « orbiter » autour de la Terre, au voisinage de celle-ci, un satellite doit être animé d'une vitesse de l'ordre de 8 000 m/s et que pour sortir du champ d'attraction terrestre (pour créer, par exemple, un nouveau satellite du Soleil) il faut communiquer à un véhicule une vitesse d'au moins 11 200 m/s (vitesse de libération ou seconde vitesse cosmique). Lorsqu'on se proposera d'envoyer vers Vénus ou Mars un astronef habité, il faudra lui communiquer une vitesse suffisante, de l'ordre de 20 000 m/s, par exemple, pour qu'il puisse accomplir sa mission dans un temps compatible avec le rythme de vie humain.

Pour simplifier, on a coutume de faire l'hypothèse que la vitesse nécessaire au mobile pour accomplir sa mission lui est communiquée instantanément, et qu'ensuite sa trajectoire est purement balistique, c'est-à-dire sous l'effet de la seule pesanteur. La vitesse ainsi déterminée est dite « vitesse caractéristique ». Une mission compliquée peut elle-même être décomposée en un certain nombre de missions élémentaires ayant chacune sa vitesse caractéristique; la vitesse caractéristique de l'ensemble sera alors naturellement la somme vectorielle de celles-ci.

Les vitesses qui viennent d'être évoquées sont d'un ordre de grandeur auquel nous ne sommes pas accoutumés parce que nous nous déplaçons toujours dans l'atmosphère terrestre, et que les vitesses que les avions les plus rapides peuvent atteindre de cette façon sont limitées par la résistance opposée par l'air à leur mouvement. Si l'on se propose de se mouvoir en dehors de l'atmosphère à l'aide d'un moyen de propulsion n'empruntant rien au milieu ambiant, de telles limitations n'existent plus. Par contre, il faut effectivement disposer d'un moyen de propulsion absolument indépendant du milieu extérieur. Ce moyen de propulsion est la fusée. On verra plus loin qu'il n'est pas, *a priori*, théoriquement impossible d'envisager d'autres méthodes pour se mouvoir dans l'espace, mais que, pratiquement, on ne sait le faire, au moins à l'heure actuelle, qu'à l'aide de fusées.

un peu de mécanique...

Qu'il s'agisse de la marche, du déplacement d'une automobile, d'un avion ou d'une fusée, un même mécanisme fondamental est mis en jeu, et il est impropre, en particulier, de parler de « propulsion par réaction » à propos d'une fusée ou d'un turboréacteur, comme on le fait couramment, dans le dessein de différencier ces modes de propulsion modernes de ceux qui ont été jusqu'ici pratiqués.

Et les mêmes lois s'appliquent : une conséquence directe de la formule fondamentale de la mécanique ($\vec{F} = m\vec{\gamma}$) peut s'énoncer facilement en introduisant la quantité de mouvement $m\vec{v}$ d'un corps et la grandeur analogue relative à un ensemble de masses (v. tome II, chap. *La mécanique à l'heure spatiale*) : les mouvements, même désordonnés, d'un ensemble de masses parfaitement isolé de toute action en provenance du milieu extérieur s'effectuent de façon telle que la somme de leurs quantités de mouvements par rapport à un système de référence galiléen demeure constante :

$$\Sigma M\vec{v} = \text{constante}.$$

Pour un ensemble de masse (M_0, M_1, M_2,...) au repos dans un tel système de référence, elle s'écrit :

$\Sigma M\vec{v} = \vec{0}$: *une masse quelconque de ce système isolé ne pourra acquérir une vitesse non nulle que dans la mesure où d'autres masses du système auront été mises en mouvement dans la direction opposée* (afin que la quantité de mouvement totale reste nulle).

Regardons un sauteur au moment précis où il prend « appel » sur le sol; l'ensemble du sauteur et de la Terre forme un système que nous pouvons supposer parfaitement isolé de l'extérieur (l'attraction universelle des planètes ne jouant aucun rôle ici!), et la quantité de mouvement de cet ensemble par rapport à un repère galiléen doit rester nulle (du moins si l'on néglige l'influence de la rotation de la Terre); cela suppose que si le sauteur acquiert la vitesse $\vec{V}$ selon la verticale, la Terre acquiert la vitesse $\vec{v}$ dans la direction opposée, de façon telle que si m est la masse de l'athlète et M celle de la Terre, on ait constamment

$$m\vec{V} + M\vec{v} = \vec{0}.$$

Il en résulte que la Terre acquiert la vitesse

$$v = V\frac{m}{M}$$

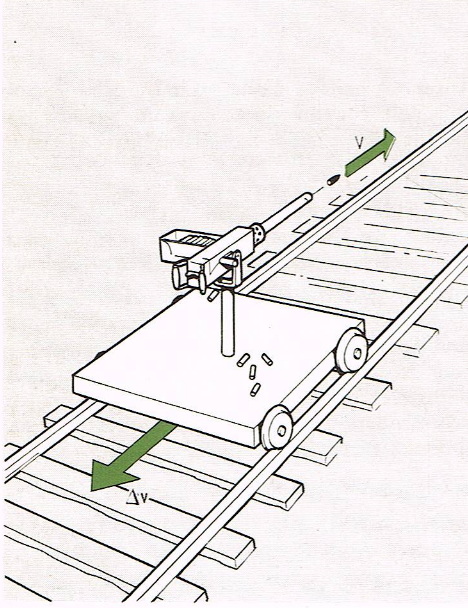
évidemment très petite et parfaitement négligeable.

Phot. U. S. I. S.

Quoi qu'il en soit, ce sauteur n'a pu s'élever que :
1° parce qu'il était capable de fournir une certaine énergie;
2° parce qu'il disposait d'une masse autre que la sienne propre à déplacer.

De la même façon, une automobile, un piéton ou un cycliste utilisent la masse de la Terre pour se déplacer. Un avion de même, qu'il soit muni d'une hélice ou d'un turboréacteur, utilise l'air ambiant, et le système à considérer est alors constitué par l'ensemble avion-atmosphère environnante, le rôle de la machine étant de fournir de l'énergie pour mettre en mouvement cette masse d'air, et par voie de conséquence l'avion lui-même, vers la direction opposée.

Supposons que l'on monte une mitrailleuse sur un wagonnet pouvant se déplacer sur des rails. Ce système n'est pas en toute rigueur parfaitement isolé puisqu'il est soumis à l'action de la pesanteur ainsi qu'à la résistance opposée par l'air et par les rails au déplacement du wagonnet. Faisons en sorte, toutefois, de tirer parallèlement à la voie, afin que l'action due à la présence des rails soit simplement une résistance de frottement, et en première approximation admettons que l'ensemble de ces actions extérieures puisse être négligé. L'ensemble du wagonnet, de la mitrailleuse et des balles tirées ou non constitue pratiquement un système isolé, et le théorème de la quantité de mouvement peut lui être appliqué. Si l'ensemble est au repos à l'instant initial, sa quantité de mouvement restera nulle quelles que soient les positions respectives des balles et du wagonnet puisque aucune action extérieure ne s'exerce sur le système. Chaque fois que la mitrailleuse tirera une balle à la vitesse V, il en résultera donc un accroissement de vitesse Δv du wagonnet dans la direction opposée à celle selon laquelle aura été tirée la balle.



Principe de la propulsion par réaction.

Pour réaliser ce mouvement, il faut disposer à bord du wagonnet :
— de masses à mettre en mouvement (les balles);
— d'une source d'énergie pour accélérer celles-ci (la poudre contenue dans les cartouches).

Un tel dispositif peut, en toute rigueur, être qualifié de « fusée » puisque le mouvement du wagonnet ne doit rien au milieu extérieur. Nous dirons que la particularité de la fusée est de disposer à bord même du véhicule d'une source d'énergie et d'une source de masse. Le sauteur, l'automobile, l'avion disposent d'une source d'énergie (muscles ou moteur), mais ils ne peuvent se déplacer qu'en mettant en mouvement une masse extérieure, et leur fonctionnement suppose donc la présence de celle-ci. **La fusée est le seul moyen de propulsion capable de fonctionner en l'absence de tout système de masses extérieures.**

Si les vitesses à communiquer aux astronefs pouvaient l'être en prenant appui sur la Terre, il ne serait pas nécessaire de faire appel à une fusée comme moyen de propulsion, et, de plus, cette manière de faire pourrait être moins coûteuse si l'on compare l'énergie cinétique finalement communiquée au véhicule à l'énergie totale qu'il aurait fallu dépenser.

Cette énergie se répartit évidemment, sous forme d'énergie cinétique, au profit de toutes les masses mises en mouvement dans l'opération. Supposons que l'on se propose de lancer un satellite à proximité de la Terre, il faudra lui fournir une énergie cinétique $E = 1/2\ m\ (8\,000)^2$ joules : si l'on opère à l'aide d'une fusée, il faudra dépenser une énergie totale qui sera la somme de E et de l'énergie cinétique communiquée aux corps éjectés : la dépense totale sera ainsi sensiblement supérieure à E. Si l'on savait utiliser la Terre comme source de masse, par contre, on communiquerait à celle-ci une vitesse et une énergie infiniment petites, et au total on ne dépenserait que E.

Malheureusement, on ne voit guère que deux possibilités théoriques pour opérer de cette façon, et l'on ne sait mettre en œuvre ni l'une ni l'autre.

La première est celle du canon de Jules Verne. Indépendamment du fait qu'il n'est pas possible de faire traverser les basses couches atmosphériques à un corps lancé depuis le sol à 8 000 ou 11 000 m/s sans le volatiliser, on ne connaît aucun moyen pour accélérer un corps à de telles vitesses en agissant simplement sur lui par action de contact.

La seconde est, *a priori*, beaucoup plus séduisante, mais toujours aussi impraticable, au moins dans l'immédiat. Si l'on savait agir à distance sur la Terre en créant, par exemple depuis le véhicule que l'on se propose d'accélérer, un champ de gravitation artificielle, il suffirait alors de disposer à bord de celui-ci d'une source d'énergie capable d'entretenir ce champ et d'utiliser la Terre ou une autre planète comme source de masse. Malheureusement, on ne sait créer à l'heure actuelle que des champs électriques et magnétiques agissant à courte distance, et un tel mode de propulsion ne peut non plus être envisagé.

mécanique de la fusée

Si les mises en mouvement de masses, dans un système isolé, se font de façon telle que le théorème de la quantité de mouvement soit respecté, il n'empêche que chaque masse considérée individuellement se déplace conformément à la loi fondamentale de la dynamique : $\vec{F} = m\vec{\gamma}$, F étant la force appliquée à cette masse, $\vec{\gamma}$ l'accélération, c'est-à-dire l'accroissement de vitesse qu'elle prend par unité de temps sous l'action de cette force : $\vec{\gamma} = \frac{d\vec{v}}{dt}$.

Revenons à notre wagonnet et à sa mitrailleuse : à un accroissement de vitesse $\Delta v = \frac{mV}{M}$ **par balle** (m étant la masse de la balle, V sa vitesse d'éjection et M la masse de l'ensemble mobile) correspond un accroissement de vitesse par unité de temps égal à $\frac{Nm}{M}$ V, N étant le nombre de balles tirées par **unité de temps.**

La force appliquée à ce wagonnet est donc égale à

$$F = M \cdot \frac{dv}{dt} = (Nm).V$$

c'est-à-dire au produit de la quantité de masse (Nm) éjectée par unité de temps par la vitesse d'éjection V. Nous appellerons cette force la « poussée ».

En général, on caractérise une fusée par sa poussée :

$$\mathbf{F} = q\, V_e,$$

q étant le débit de masse (par seconde), V_e la vitesse d'éjection.

On a coutume, cependant, de caractériser un moteur par sa puissance, c'est-à-dire par la quantité de travail qu'il peut fournir par unité de temps. Pourquoi ne pas faire de même lorsqu'il s'agit d'une fusée?

La notion de puissance prend toute sa signification lorsqu'il s'agit d'une machine fixe, telle une turbine à vapeur, le but étant en définitive de fournir de l'énergie directement utilisable. Mais lorsqu'on se propose de mettre un mobile en mouvement, cette notion présente beaucoup moins d'intérêt puisque l'énergie fournie se répartit au profit de toutes les masses mises en mouvement et non de la seule masse du mobile qu'on se propose d'accélérer.

Une fusée n'étant pas destinée à créer de l'énergie, mais bien du mouvement, il est donc légitime de la caractériser par une grandeur directement liée au mouvement. Celle-ci est évidemment la poussée.

Si l'on se propose de réaliser un moteur d'une certaine poussée, on dispose *a priori* d'une infinité de solutions, si on peut effectivement choisir arbitrairement les valeurs de q et de V_e dont le produit est la poussée. Il est clair, toutefois, que si l'on opte pour éjecter à faible vitesse et à grand débit, la masse de lest à transporter ainsi que le poids total du véhicule pour obtenir l'accroissement de vitesse voulu seront très grands. A l'inverse, le choix d'une vitesse d'éjection plus élevée permettra de transporter moins de lest; c'est-à-dire, toutes choses égales par ailleurs, de diminuer le poids total du véhicule. Nous verrons plus loin que pour plusieurs raisons on n'est maître du choix de la vitesse que dans une certaine mesure, et que l'on peut souvent mettre en évidence l'existence pratique d'une valeur optimale de celle-ci. Retenons, toutefois, en première approximation, **que la recherche de moyens permettant de réaliser des vitesses d'éjection élevées est un souci constant pour les spécialistes des problèmes de propulsion.**

L'intérêt de cette recherche apparaît d'ailleurs clairement si l'on se réfère à l'équation de base de la propulsion par fusée, qui donne, en l'absence de tout champ de force et de toute résistance extérieure, l'accroissement de vitesse correspondant à l'éjection d'une certaine quantité de lest ($M_0 — M_1$) à la vitesse V_e par rapport à la fusée :

$$V_1 - V_0 = V_e \log. \frac{M_0}{M_1}.$$

Notons au passage que cette équation peut d'ailleurs être aisément déduite de la loi fondamentale de la dynamique et de l'expression de la poussée.

Essayons de faire apparaître quelques ordres de grandeur. Supposons, par exemple, qu'on se propose de communiquer à un véhicule une vitesse de 5 000 m/s à l'aide d'une fusée, on peut calculer le rapport entre la masse initiale et la masse finale (M_0/M_1) en fonction de la vitesse d'éjection :

V_e	1 000	2 500	5 000	10 000	100 000 m/s
M_0/M_1	149	7,4	2,72	1,64	1,05

Comme on peut en juger, la masse de lest à transporter décroît très rapidement lorsque la vitesse d'éjection augmente. Il n'en résulte pas nécessairement, toutefois, pour la même valeur de $V_1 — V_0$, une augmentation dans les mêmes proportions de la charge utile transportée. La mise en œuvre des moyens nécessaires pour obtenir une vitesse d'éjection supérieure à quelques milliers de mètres-secondes a, en effet, pour conséquence d'augmenter le poids mort embarqué : pour éjecter des gaz à une vitesse de l'ordre de 1 000 m/s, il est suffisant d'utiliser les produits de la réaction d'un combustible et d'un oxydant : la masse à vide de la fusée est alors essentiellement constituée par la masse des réservoirs. Pour éjecter, par contre, un gaz à 10 000 m/s, il faut le chauffer dans un réacteur nucléaire; à la masse de lest que fait ressortir le rapport M_0/M_1, il convient donc d'ajouter celle, qui est loin d'être négligeable, du réacteur nucléaire. Pour éjecter des particules à une vitesse de 100 000 m/s, il est nécessaire d'avoir recours à des méthodes encore plus élaborées, et le poids mort à bord de la fusée devient nettement supérieur au poids de lest éjecté. **L'adoption d'un système à vitesse d'éjection élevée n'est donc justifiable que dans la mesure où elle n'a pas pour conséquence d'augmenter de façon inacceptable le poids mort embarqué** (ainsi d'ailleurs que la durée de la mission en nécessitant de limiter la poussée à une valeur extrêmement faible).

Il convient de noter que l'usage a prévalu de substituer une autre grandeur à la vitesse d'éjection : l'impulsion spécifique I_s, c'est-à-dire l'impulsion communiquée au véhicule par unité de poids éjectée (l'accélération intervenant dans la détermination du poids étant l'intensité de la pesanteur au niveau du sol, soit 9,81 m/s²) :

$$I_s = \frac{mV_e}{mg} = \frac{V_e}{g} \sim 0{,}1\ V_e.$$

Si V_e est exprimée dans le même système d'unité que g, l'impulsion spécifique s'exprime en secondes. (Bien que cette grandeur soit dépourvue de toute signification physique ou mécanique, son emploi s'est cependant généralisé, probablement, entre autres, parce qu'elle s'exprime par le même nombre en unités métriques et britanniques, qui ont en commun l'unité de temps.)

le rapport de masse

A défaut de savoir réaliser des vitesses d'éjection élevées, l'équation fondamentale de la propulsion par fusée nous montre qu'on peut effectuer la même mission avec une vitesse d'éjection plus faible et un rapport entre masse initiale et masse finale plus grand, c'est-à-dire en transportant une quantité de masse à éjecter plus importante. Cependant, on ne peut pas augmenter indéfiniment ce rapport de masse :
— en raison des dimensions inacceptables qui en résulteraient;
— parce que, de toute façon, à un niveau de possibilités technologiques données correspond une valeur limite de ce rapport.

En effet, la masse initiale de la fusée peut être approximativement représentée par la somme de trois termes :
— la masse utile (la masse du satellite proprement dit par exemple), que nous appellerons M_u;
— la masse de lest à éjecter (celle du combustible et du comburant dans les fusées classiques), que nous appellerons M_p;
— la masse de la fusée à vide qui, dans les fusées à source d'énergie chimique, est approximativement proportionnelle à la masse totale de combustible et de comburant, que nous appellerons M_s; $M_s = kM_p$.

La masse initiale est donc égale à $M_0 = M_p\ (1 + k) + M_u$.

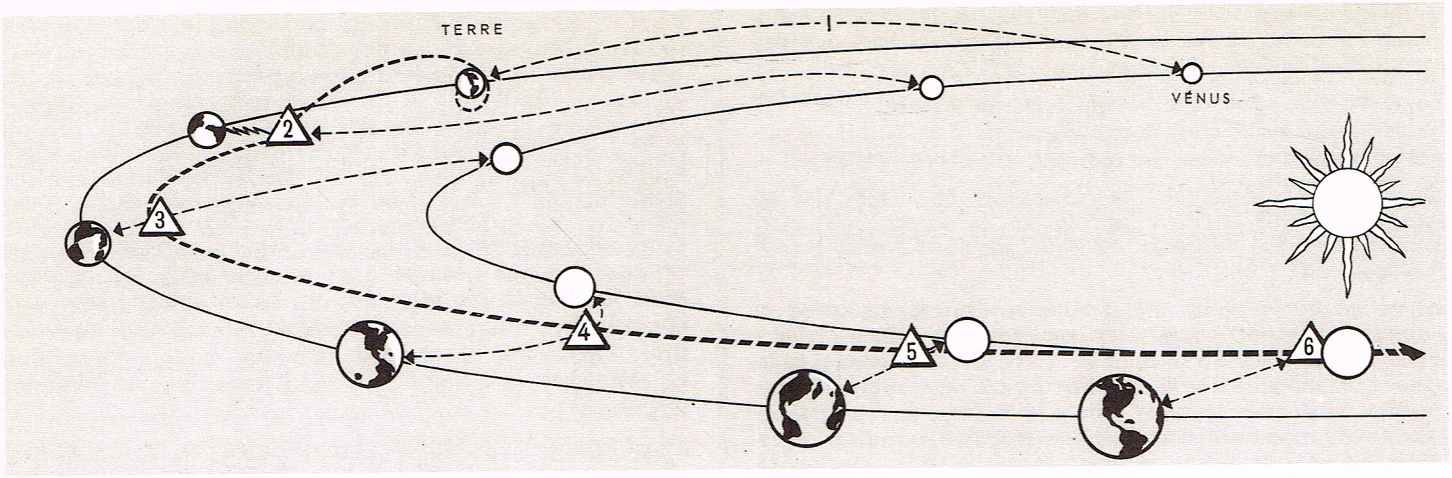
Quant à la masse finale, elle est égale à $M_1 = kM_p + M_u$.

Considérons maintenant l'expression du rapport :

$$\frac{M_0}{M_1} = \frac{(1 + k)\ M_p + M_u}{kM_p + M_u}.$$

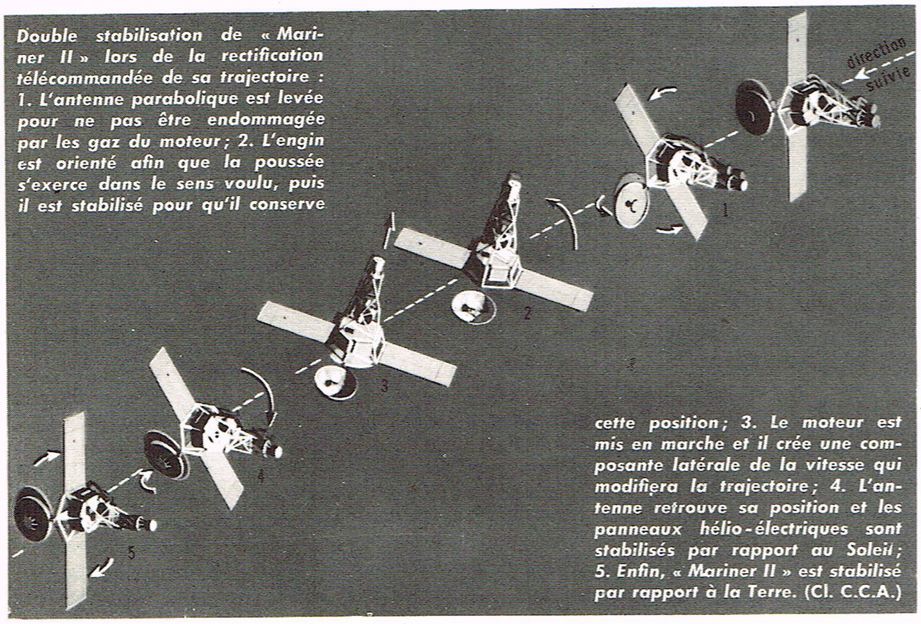
Il est clair que si M_p devient très grand, c'est-à-dire si M_u peut être négligé, la valeur limite de $\frac{M_0}{M_1}$ est égale à $\frac{1 + k}{k}$.

Si l'on sait faire une structure extrêmement légère, k pourra



Positions successives
de la Terre, de Vénus et de « Mariner II » (Δ).

1. Le 27 août 1962, lors du lancement de l'engin;
2. Le 4 septembre, lors de la rectification de sa trajectoire;
3. Le 1er octobre;
4. Le 1er novembre;
5. Le 1er décembre;
6. Le 14 décembre, lors du survol de Vénus par la sonde spatiale.



Double stabilisation de « Mariner II » lors de la rectification télécommandée de sa trajectoire : 1. L'antenne parabolique est levée pour ne pas être endommagée par les gaz du moteur; 2. L'engin est orienté afin que la poussée s'exerce dans le sens voulu, puis il est stabilisé pour qu'il conserve cette position; 3. Le moteur est mis en marche et il crée une composante latérale de la vitesse qui modifiera la trajectoire; 4. L'antenne retrouve sa position et les panneaux hélio-électriques sont stabilisés par rapport au Soleil; 5. Enfin, « Mariner II » est stabilisé par rapport à la Terre. (Cl. C.C.A.)

Pour accomplir une mission telle que la mise sur orbite d'un satellite, ou *a fortiori* une mission plus ambitieuse, on a généralement recours à l'emploi de véhicules lanceurs constitués de plusieurs fusées superposées les unes aux autres. L'allumage de chacune est provoqué à l'instant jugé optimal pour la mission considérée.

Considérons l'exemple de « Mariner II » : le véhicule lanceur était un « Atlas-Agena B », composé d'un premier étage « Atlas » et d'un second étage « Agena B », la charge utile étant constituée par « Mariner » lui-même. L'opération s'est effectuée en trois temps :

— Mise sur une orbite terrestre (fonctionnement total d'« Atlas », puis largage de celui-ci, fonctionnement partiel d'« Agena »);

— Réallumage d'« Agena » et mise sur une trajectoire de transfert vers Vénus;

— A mi-course vers Vénus, manœuvre à l'aide d'une petite fusée transportée par « Mariner » pour corriger la trajectoire de celui-ci.

Le départ depuis le sol s'effectuant verticalement, il était nécessaire que la poussée d'« Atlas » fût supérieure au poids total du véhicule. Par contre, aucune limitation de ce genre n'intervenait en ce qui concerne la poussée d'« Agena » puisque son premier allumage eut lieu alors qu'il était à peu près à l'horizontale et son second lorsqu'il était déjà satellisé. Si ΔV a été l'accroissement de vitesse communiqué par « Agena » à « Mariner » ce pouvait donc être aussi bien en faisant fonctionner pendant une longue durée une fusée de faible poussée ou pendant peu de temps une fusée de forte poussée. Il est probable que, dans ce cas précis, la poussée et le temps de fonctionnement d'« Agena » correspondaient, au moins de façon approchée, aux conditions optimales, mais d'autres configurations pourraient être plus avantageuses pour d'autres missions.

On peut donc être conduit à construire des fusées de poussées extrêmement différentes, allant de plusieurs centaines de tonnes ou plus, à quelques centaines de grammes ou moins, selon le but que l'on se propose.

être petit, mais il ne pourra jamais être nul, et M_0/M_1 est nécessairement limité.

Le coefficient de proportionnalité *k* dépend :
— de la densité du comburant et du combustible;
— de l'arrangement général du véhicule et de sa taille;
— des caractéristiques mécaniques des matériaux utilisés, et, d'une façon générale, du degré d' « élaboration » des solutions technologiques adoptées.

Pratiquement, pour la plupart des véhicules au sujet desquels on dispose d'informations suffisantes, la valeur de *k* est comprise entre 0,04 et 0,12.

En résumé, le rapport de masse étant limité par des sujétions d'ordre technologique, ainsi que la vitesse d'éjection, on pourrait en déduire que les missions spatiales que l'on peut envisager d'entreprendre sont limitées à celles qui correspondent à une vitesse caractéristique inférieure à la valeur limite définie par

$$\Delta V_{limite} = V_{e\ max} \log (M_0/M_1)_{max}.$$

Dans ces conditions, on serait actuellement à peine capable de placer un tout petit satellite sur orbite! Il existe donc une méthode permettant de lever, ou tout au moins de reculer, cette limitation. Cette méthode consiste à utiliser une fusée à plusieurs étages.

pourquoi plusieurs étages ?

Pour quelle raison le lancement de « Mariner », par exemple, qui n'aurait pu être effectué à l'aide d'une seule fusée, est-il devenu possible en en superposant deux. On peut le concevoir aisément de façon quasi empirique, à défaut du calcul mathématique rigoureux, qui donnerait pourtant facilement la manière optimale d'opérer dans chaque cas.

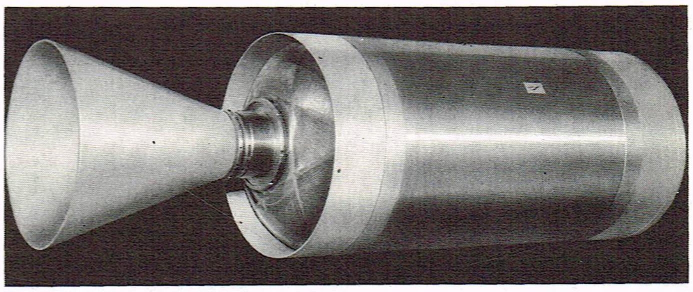
Une locomotive à vapeur doit se ravitailler en eau et en charbon à des postes fixes échelonnés à une certaine distance le long du trajet parcouru par le train. Supposons qu'on se propose de faire franchir à un train une chaîne de montagnes très élevée et que la distance à parcourir soit très longue, et de toute façon supérieure à celle qui est prévue normalement entre les postes de ravitaillement, tant est faible la pente maximale possible pour un tel convoi. Supposons, en outre, que ces postes intermédiaires aient leurs réservoirs vides et qu'il faille que notre train transporte depuis le bas la réserve de charbon et d'eau qui lui est nécessaire pour atteindre le sommet. Au fur et à mesure du trajet, les « containers » de charbon et les citernes d'eau s'épuiseront. S'il est possible d'abandonner ceux-ci dès qu'ils seront vides, cette manière de faire aura été certainement plus coûteuse que la manière usuelle, mais on aura toutefois limité l'excédent d'énergie à fournir au strict minimum. Par contre, si l'on ne peut abandonner les wagons vides, le travail nécessaire pour élever ceux-ci de quelques milliers de mètres devra être fourni par la locomotive, et un surcroît de charbon et d'eau devra être prévu pour fournir cette énergie. Les wagons correspondant à ce supplément devront eux-mêmes être acheminés et demanderont donc une quantité supplémentaire de charbon, d'eau,... qui nécessitera de nouveaux wagons supplémentaires, etc.

Dans le premier cas, on réduira au minimum l'énergie excédentaire à fournir en multipliant le nombre de containers et en abandonnant ceux-ci au fur et à mesure qu'ils seront vides; dans le second, le poids mort transporté sera minimal, toutes choses égales par ailleurs, lorsqu'on n'utilisera qu'un seul container pour le charbon et qu'une seule citerne pour l'eau. Il existe, bien entendu, des solutions intermédiaires : ce sont celles où l'on s'autorisera à abandonner les containers vides en un certain nombre de points le long du trajet.

Une fusée quittant le sol ne peut disposer, dans l'état actuel des choses, d'aucun moyen de se ravitailler en cours de route : elle doit donc transporter tout ce qui lui est nécessaire, en particulier les réservoirs dans lesquels est contenu le lest qui sera éjecté (combustible et comburant dans une fusée classique, par exemple). Si l'on savait abandonner à chaque instant la masse de réservoirs correspondant à la quantité de lest éjecté, on opérerait à peu près de la même façon que le train abandonnant au fur et à mesure ses petits wagons de charbon vides, et la mission serait accomplie avec la dépense d'énergie minimale. Au contraire, si l'on s'impose de transporter pendant toute la durée de celle-ci la masse de réservoirs qui n'est plus d'aucune utilité, on sera conduit à une fusée de dimensions beaucoup plus importantes.

Il n'est évidemment pas possible d'éjecter simultanément du lest et la partie de réservoirs correspondante. Mais à défaut de savoir opérer de cette façon idéale, on conçoit qu'il est possible d'abandonner en une ou plusieurs fois les charges qu'il est devenu inutile de transporter, comme le train abandonnant ses containers en un certain nombre de points. **On a ainsi réalisé une fusée à étages.**

Il convient toutefois de noter que la modification des vitesses d'un système de masses peut se faire par des procédés purement passifs, dans la mesure où ce système n'est pas isolé du milieu extérieur. C'est le cas, par exemple, si la masse que l'on se propose de mettre en mouvement est placée dans le rayonnement solaire. Si l'on déploie une surface suffisamment importante perpendiculairement aux rayons, les photons captés ou réfléchis par cette voile communiquent au véhicule une certaine quantité de mouvement : un photon d'énergie $h\nu$ a une masse équivalente $= h\nu/c^2$, d'où, puisqu'il se déplace à la vitesse de la lumière c, une quantité de mouvement $= h\nu/c$. S'il est absorbé par la voile (corps noir), cette quantité de mouvement sera communiquée au véhicule. S'il est parfaitement réfléchi, c'est la quantité $2\ h\nu/c$ qui sera transmise à celui-ci.



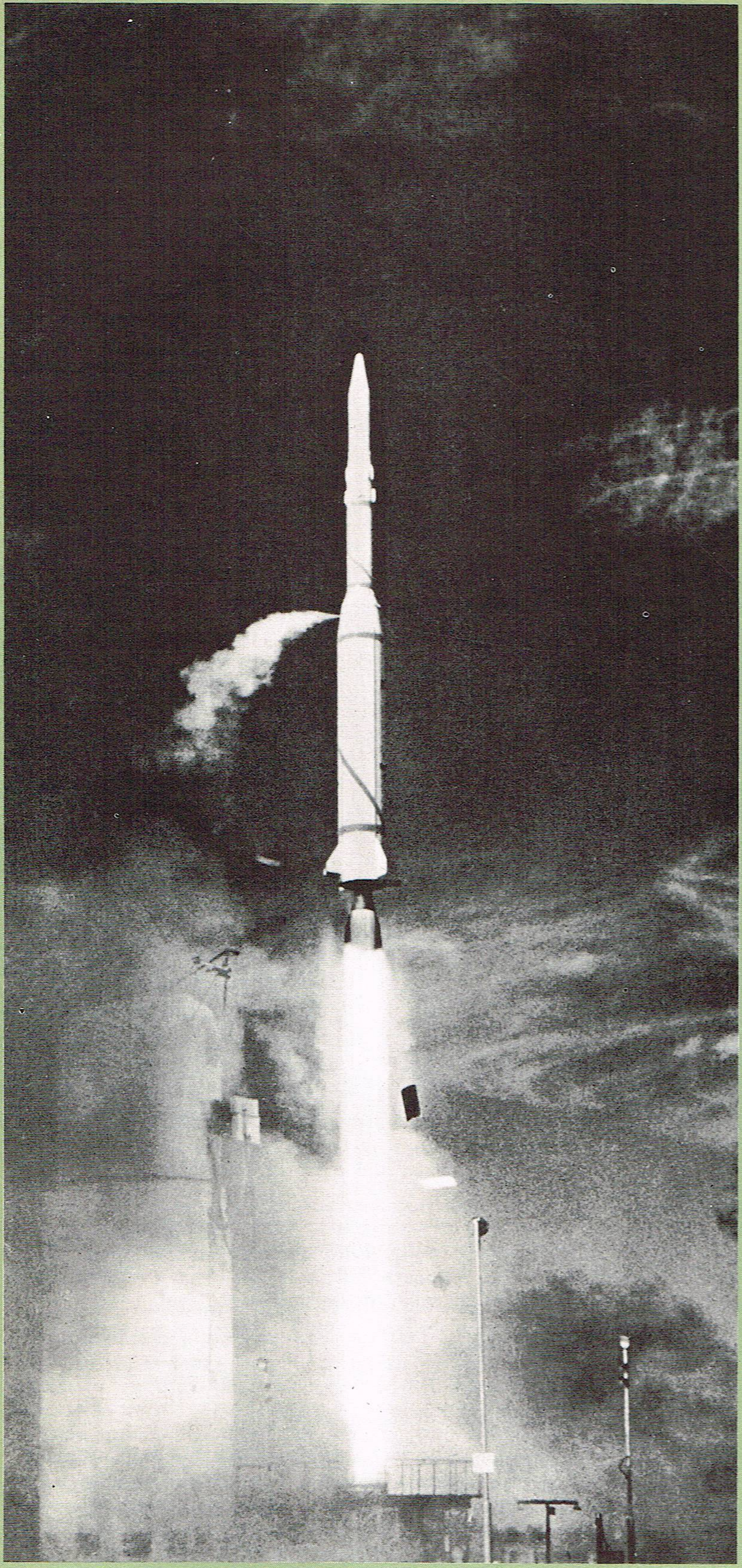
Troisième étage du lance-satellite français « Diamant », à propulsion solide.

Malheureusement, un tel procédé conduit également à des poussées extrêmement faibles, la pression obtenue sur la voile, même dans le cas d'une réflexion parfaite, étant, au voisinage de la Terre, de l'ordre de 10^{-10} atmosphères. Il n'est certes pas exclu que dans le futur, et pour des applications bien particulières, un tel mode de propulsion soit envisageable; toutefois, pour les missions spatiales réalisées aujourd'hui, aussi bien que pour celles que l'on envisage dans un avenir proche, l'extrême petitesse de la poussée ainsi obtenue rend l'utilisation d'un tel système absolument rédhibitoire.

Il n'est pas impossible, cependant, de penser utiliser l'énergie solaire, mais d'une façon plus efficace qu'en se limitant à bénéficier de la pression de radiation.

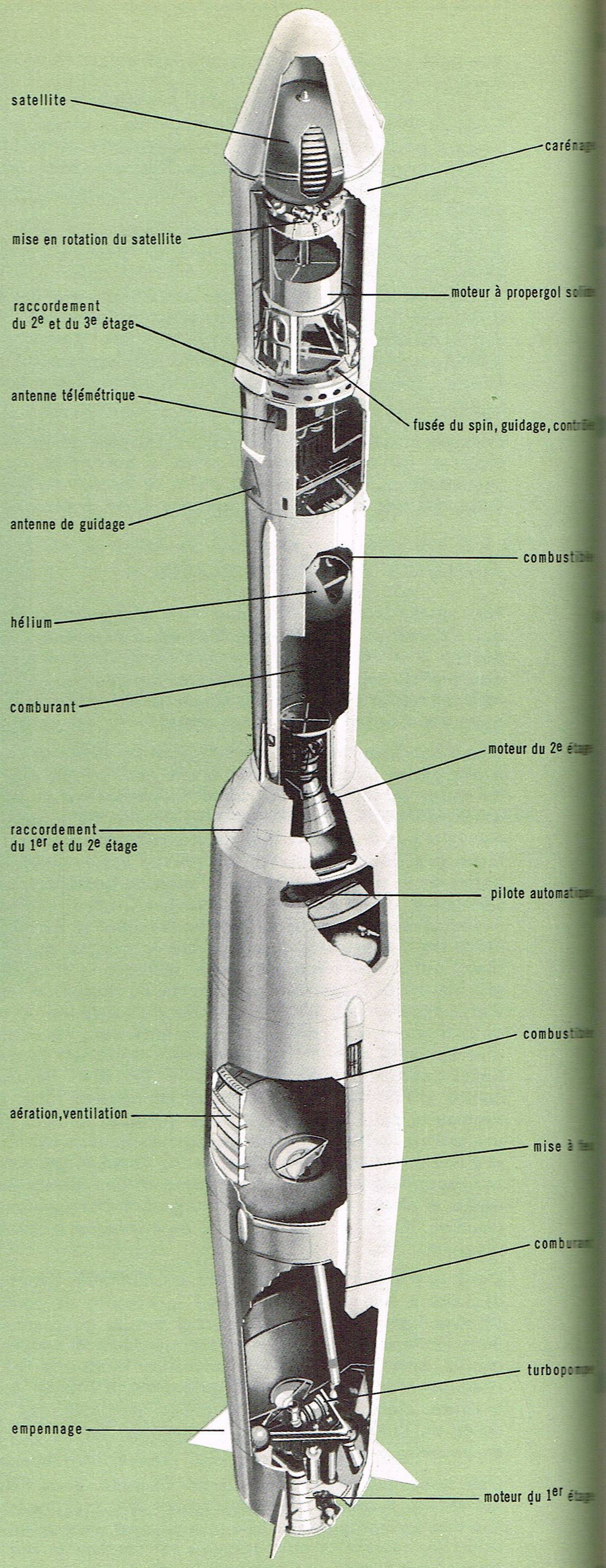
Nous verrons au chapitre consacré aux satellites qu'il existe différentes méthodes pour transformer à bord d'un véhicule spatial l'énergie solaire en énergie électrique ou mécanique. Les moyens technologiques actuels ne permettent pas, toutefois, d'envisager ce mode de propulsion dans l'immédiat.

Phot. Larousse.



La première fusée " Diamant " s'élève au-dessus de sa plate-forme de lancement à Hammaguir, dans le Sud algérien.

FUSÉE «THOR DELTA» A TROIS ÉTAGES



Phot. C. C. A.

Pour fixer les idées, supposons que l'on se propose de placer sur une orbite proche de la Terre une charge d'une tonne. Compte tenu de la résistance de l'air et du travail effectué contre la pesanteur durant le fonctionnement de la fusée, la vitesse caractéristique à lui communiquer est de l'ordre de 10 000 m/s.

Si l'on dispose de moyens technologiques permettant de réaliser pour chaque étage un rapport $k = 0{,}08$ entre la masse de la fusée à vide et la masse de lest, et si la vitesse d'éjection pour chaque étage est supposée être égale à 2 700 m/s, la masse totale du véhicule sera :
— en un seul étage : infini (pas de solution).
— en deux étages : 126 t.
— en trois étages : 77 t.
— en quatre étages : 68 t.

Ainsi qu'on l'a déjà vu plus haut, il n'y a pas de solution en un seul étage. Toutefois, lorsqu'on augmente le nombre de ceux-ci ce gain va rapidement en diminuant, et, compte tenu des difficultés inhérentes à l'augmentation du nombre des étages et des causes de mauvais fonctionnement qui peuvent en résulter, il est pratiquement sans intérêt, pour la mission que nous avons envisagée et les hypothèses que nous avons faites, d'aller au-delà de deux ou trois étages. En règle générale, le choix du nombre d'étages se fait en essayant d'évaluer le prix de revient global d'une mission ou de l'ensemble des missions pour lesquelles on utilisera le lanceur étudié, plutôt que dans le dessein de minimiser autant que possible le poids au départ de celui-ci, lequel n'est, en définitive, qu'un des éléments et non nécessairement le plus important, qui conditionnent le prix de revient d'une mission.

comment se procurer l'énergie nécessaire ?...

Il nous faut retenir des paragraphes qui précèdent deux notions fondamentales :

— Les paramètres qui régissent le mouvement d'une fusée sont, indépendamment de ceux qui sont imposés par le milieu extérieur, la masse instantanée de celle-ci et sa poussée;
— La masse de lest à transporter pour effectuer une mission donnée sera d'autant plus faible que la vitesse à laquelle on l'éjectera sera plus élevée.

Toutes choses égales par ailleurs, la vitesse d'éjection peut donc être considérée comme un indice de qualité, et la recherche de moyens pour l'améliorer un souci légitime.

En fait, obtenir une vitesse d'éjection élevée n'est pas en soi un problème particulièrement difficile à résoudre : un émetteur de rayons γ, par exemple, permet d'éjecter des particules à la vitesse de la lumière; ces particules sont des photons, de masse rigoureusement nulle, mais possédant néanmoins une faible quantité de mouvement; toutefois, la poussée résultant de l'éjection de quantités énormes de photons est à peine sensible. Accélérer une masse infinitésimale n'est donc pas une opération intrinsèquement difficile; malheureusement, si le but que l'on se propose est d'obtenir de cette façon un effet propulsif, elle est sans intérêt.

Le problème consiste donc à obtenir simultanément une poussée d'une valeur suffisante pour accomplir la mission que l'on désire effectuer et une vitesse d'éjection élevée; c'est-à-dire de disposer à bord de la fusée d'une source d'énergie capable de satisfaire à ces conditions.

Pour fixer les ordres de grandeur des énergies nécessaires pour l'éjection, nous donnons ci-dessous l'énergie qu'il faut communiquer à l'unité de masse éjectée en fonction de sa vitesse :

V_e	2 500	5 000	10 000	100 000 m/s;
W	3,125	12,5	50	5 000 kJ/g.

L'énergie totale à fournir est évidemment égale au produit de la masse totale de lest par la valeur de W correspondant à la vitesse d'éjection désirée, affecté des divers rendements liés aux transformations que subit la nature de l'énergie à partir de la forme sous laquelle elle se trouve initialement disponible à bord de la fusée. Si l'on effectue le calcul dans des cas précis, on est conduit à des résultats impressionnants : la quantité d'énergie qui doit être emmagasinée à bord de la fusée « Atlas », par exemple, est de l'ordre de 10^9 kJ, soit sensiblement plus que la production horaire d'une des grosses unités de 250 MW équipant les centrales thermiques les plus récentes, fonctionnant à pleine puissance.

On conçoit qu'il ne peut être question de stocker de telles quantités d'énergie à l'aide des moyens auxquels on a généralement recours au sol : batteries d'accumulateurs, volants d'inertie, etc., les masses auxquelles de telles solutions conduiraient étant inacceptables.

Pratiquement, les deux seules sources d'énergie suffisamment concentrées qui puissent être utilisées à cette fin sont :
— **les sources d'ordre chimique** dans lesquelles l'énergie est libérée par la mise en œuvre d'une réaction chimique fortement exothermique;
— **les sources d'ordre nucléaire** qui, *à priori*, peuvent être de trois types différents :
a) sources à radio-isotopes à base d'éléments radio-actifs émettant des rayonnements α, β et γ, dont l'énergie peut être transformée en chaleur, puis en électricité, par exemple. (Pour des poussées et des puissances instantanées faibles, une telle solution n'est pas toutefois à rejeter *a priori*, bien que l'emploi de radio-isotopes pour la propulsion nécessiterait la mise en œuvre de moyens de production extrêmement coûteux.)
b) réacteur à fission;
c) réacteur utilisant l'énergie de fusion de deux atomes de deutérium en un atome d'hélium.

On sait toutefois que cette source d'énergie, beaucoup plus puissante même que l'énergie de fission, ne peut être mise en œuvre — au moins actuellement — que dans des conditions qui ne permettent pas de contrôler sa libération (bombe H).

En fait, seules les sources d'énergie chimique (à propergols liquides ou solides) et les réacteurs nucléaires utilisant la fission de l'atome d'uranium ou d'un autre élément fissile sont donc susceptibles de fournir à bord d'une fusée l'énergie nécessaire à assurer la propulsion de celle-ci.

... et la transformer en énergie cinétique ?

Quelle que soit la forme première sous laquelle se trouve l'énergie, celle-ci doit être convertie en énergie cinétique utile communiquée à la masse de lest dont l'éjection aura pour effet de propulser la fusée. Il n'existe pratiquement que deux méthodes fondamentalement différentes pour effectuer cette transformation, lesquelles conduisent d'ailleurs à des systèmes de propulsion n'ayant de commun que le fait de satisfaire à la définition de la fusée comme un moyen de propulsion disposant de sa propre source de masse. On peut, en effet, soit agir directement par action de contact sur ces masses (c'est ce que l'on réalise en détendant un gaz), soit les placer dans un champ électrique ou électromagnétique après les avoir rendues porteuses d'une charge électrique.

Ces deux méthodes serviront de base à notre classification des différents types de fusées; nous appelons celles du premier type « fusées thermiques », celles du second type « fusées électriques ».

les fusées thermiques

Toutes les fusées de ce genre fonctionnent selon le processus suivant :

Si l'on utilise une source d'énergie chimique, c'est-à-dire en général la réaction d'un combustible et d'un comburant, ce sont les gaz produits par la combustion qui constituent la source de masse. Mais la source d'énergie peut être aussi un réacteur nucléaire ou un arc électrique. Une source de masse séparée doit alors être prévue (en général de l'hydrogène) qui s'échauffera en passant dans le réacteur nucléaire ou dans l'arc.

comment transformer l'énergie thermique en énergie cinétique ?

Dans le cas d'une fusée thermique, on dispose donc de gaz chauds que l'on se propose d'accélérer en les détendant, c'est-à-dire en les faisant passer d'une pression P_0 (la pression de combustion) à une pression P_1 inférieure à P_0. On peut montrer théoriquement, et on vérifie expérimentalement, que la façon optimale d'opérer consiste à effectuer cette détente dans une enceinte convenablement dessinée appelée « tuyère », constituée d'une partie convergente suivie d'une partie divergente, la section la plus étroite étant appelée le « col ».

Le taux de transformation en énergie cinétique de l'énergie stockée initialement sous forme thermique est évidemment fonction du rapport de détente P_0/P_1. Pour un taux de détente infini, en négligeant les phénomènes de liquéfaction et de solidification, on obtiendrait des gaz à zéro degré (absolu) et à une vitesse limite V_∞. Pratiquement, on ne peut réaliser, bien entendu, que des détentes finies, et V_e est toujours inférieur à V_∞.

Il n'est pas toujours possible, d'ailleurs, de choisir un taux de détente élevé. La pression P_0, en effet, pour des raisons généralement d'ordre mécanique, est rarement supérieure à quelques dizaines d'atmosphères, et pour des propulseurs devant fonctionner au niveau du sol, par exemple, la pression de sortie ne peut guère, pour des raisons d'ordre aérodynamique, être inférieure à la pression atmosphérique. Il en résulte que, dans le cas d'un premier étage de fusée, le taux de détente est généralement inférieur à 100/1 et la vitesse d'éjection au maximum de l'ordre de 75 p. 100 de la vitesse limite. Pour des étages supérieurs, la pression ambiante aux altitudes auxquelles ils fonctionnent étant beaucoup plus faible, il est possible d'atteindre des taux de détente de 500/1, 1 000/1, voire supérieurs, et d'obtenir ainsi des vitesses d'éjection de l'ordre de 0,85 à 0,90 de la vitesse limite.

Quoi qu'il en soit, à une détente poussée correspond toujours une tuyère d'autant plus longue, difficile à loger et lourde, et il est rare, pour ces raisons, que l'on envisage de dépasser, pour les applications actuelles, les taux de détente ci-dessus.

les propulseurs à source d'énergie chimique

Dans les propulseurs de ce genre, source d'énergie et source de masse sont confondues, et la poussée est obtenue généralement par éjection des gaz chauds provenant d'une réaction chimique fortement exothermique, pouvant être :
— une combustion, c'est-à-dire l'action d'un agent réducteur sur un agent oxydant, par exemple :

$$C + O_2 = CO_2$$
$$H_2 + 1/2\ O_2 = H_2O;$$

— la décomposition d'un corps stable à température ambiante, mais instable à température plus élevée. Par exemple, l'eau oxygénée concentrée : $H_2O_2 = H_2O + 1/2\ O_2$.

Pratiquement, seules les réactions du premier genre, qui représentent généralement des sources d'énergie plus puissantes, sont utilisées, tout au moins pour la propulsion proprement dite.

Notons, néanmoins, que la décomposition de l'eau oxygénée ou d'autres produits du même genre est souvent utilisée dans des fusées destinées à corriger la position ou l'orientation d'un satellite déjà en orbite (par exemple, dans le cas des capsules Mercury).

deux types de fusées à source d'énergie chimique

Le fonctionnement d'une fusée à source d'énergie chimique peut s'effectuer selon deux processus différents caractérisant deux types de fusées :

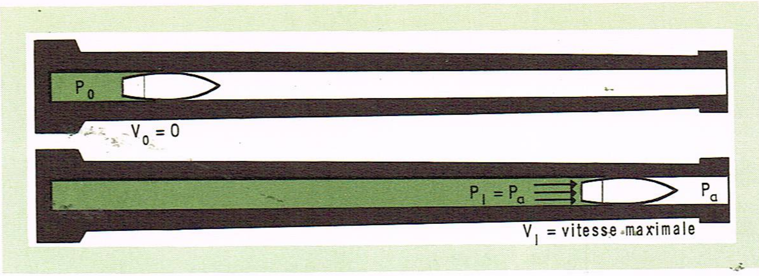
On peut provoquer l'inflammation d'un mélange de deux corps (combustible et comburant) stockés séparément et injectés à un certain débit dans une enceinte appelée « chambre de combustion ». Ce stockage est nécessairement effectué sous forme liquide; il ne peut être question, en effet, de prévoir un stockage en phase gazeuse qui conduirait, pour une même masse de produit, à des réservoirs gigantesques et extrêmement lourds, quelle que soit la pression de stockage adoptée.

On dénomme généralement ce type de fusées : **fusées à liquides.** Parmi les plus célèbres, citons le V2 allemand de la dernière guerre, les engins américains « Atlas », « Thor », « Titan », « Saturn », les engins soviétiques « T1 », « T2 », le lanceur « Europa 1 » de l'E. L. D. O.

Bien longtemps avant de savoir faire fonctionner les fusées à liquides, on a su lancer des propulseurs à poudre. On entend par **poudre ou propergol solide,** un mélange contenant les éléments nécessaires à la combustion, stable dans les conditions ambiantes de température et de pression, mais dont on peut provoquer l'inflammation par élévation de température ou par choc. Depuis un millier d'années, et depuis plus longtemps peut-être, on a su faire la poudre « noire », mélange de charbon pulvérisé et de salpêtre (nitrate de potassium). On sait aujourd'hui réaliser des poudres beaucoup plus élaborées et beaucoup plus puissantes, comme on le verra par la suite. Parmi les fusées à poudre les plus connues, citons « Polaris », « Minuteman », les fusées-sondes françaises « Bélier », « Centaure », le lanceur français « Diamant » (2ᵉ et 3ᵉ étages), etc. Aussi différents que puissent paraître au premier examen ces deux types de propulseurs, ils ne diffèrent cependant que par la manière dont l'énergie est stockée et la combustion amorcée.

à la recherche des réactions les plus énergétiques

L'énergie qui se trouve sous forme cinétique dans les gaz éjectés par une fusée de ce genre représente une certaine fraction de l'énergie libérée par la réaction mise en jeu; en aucun cas



On pourrait d'ailleurs envisager un autre type de fusée thermique fonctionnant comme la mitrailleuse qui provoque le déplacement du wagonnet (voir page 104), c'est-à-dire possédant une source d'énergie chimique (la poudre contenue dans la cartouche) et une source de masse séparée (les balles). Nous allons vérifier qu'un tel système n'est absolument pas avantageux.

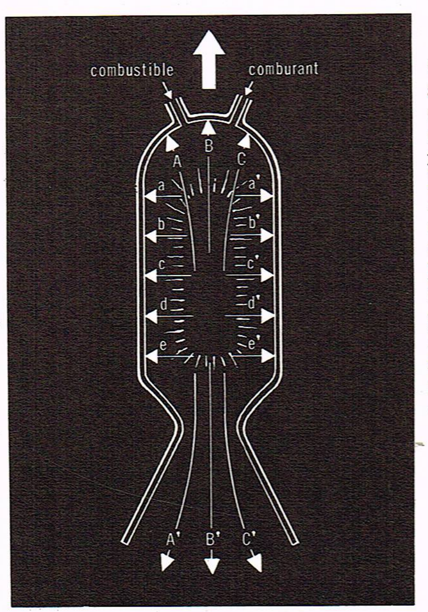
Représentons schématiquement, comme précédemment, la mitrailleuse par un tube fermé à une extrémité dont le wagonnet est solidaire. A l'instant « zéro » correspondant à la combustion supposée instantanée de la poudre, les conditions sont celles du premier cas représentées ci-dessus. La vitesse de la balle est nulle, la pression P_0 dans les gaz de combustion beaucoup plus grande que la pression atmosphérique régnant devant la balle. La force appliquée sur le culot de celle-ci étant supérieure à celle qui résulte de l'action de la pression aval, la balle va se mettre en mouvement et laisser disponible pour les gaz un volume plus grand. Au fur et à mesure de la progression de la balle dans le canon, il en résultera :

— une diminution continue de la pression des gaz ;
— l'accélération de la balle et des gaz.

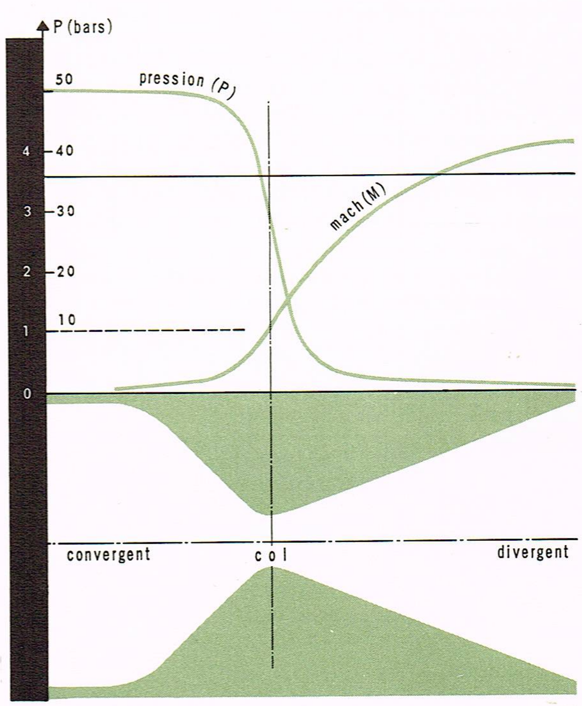
Il n'est évidemment pas possible d'éjecter la balle à une vitesse supérieure à celle des gaz dans la section immédiatement voisine du culot de la balle à l'instant où l'effet de détente cesse, pour simplifier disons dans la section où l'équilibre des pressions de part et d'autre de la balle est atteint. On comprend aussi que cette vitesse limite (pour les gaz comme pour la balle) sera d'autant plus élevée que la masse de la balle sera faible.

A la limite, si l'on remplace la masse de la balle par la même masse de poudre, on éjectera la même masse totale, mais à une vitesse beaucoup plus grande, et la poussée sera plus élevée.

En résumé, si l'on se propose d'utiliser la détente d'un gaz comme source d'énergie, à égalité de masse finalement éjectée — c'est-à-dire, en fait, transportée par la fusée — il est plus avantageux d'accélérer le gaz lui-même que d'accélérer du lest, et une fusée du type mitrailleuse ne présente de ce fait aucun intérêt.



Combustible et comburant aboutissent à la chambre de combustion de la fusée et donnent lieu à une réaction fortement exothermique. La pression des gaz en *a, b, c*... est rigoureusement équilibrée par la pression antagoniste en *a', b', c'*... En revanche, la poussée des gaz en A, B, C n'est pas compensée, puisque, en A', B', C', les molécules s'échappent vers l'extérieur, d'où une poussée résultante qui projette l'engin dans le sens de la flèche.



Pressions et vitesses dans une tuyère propulsive.

elle ne peut être supérieure à celle-ci. Pour un débit masse m on peut donc écrire la relation suivante :

$$1/2\ m\ V_e^2 = \eta\ m\ Q,$$

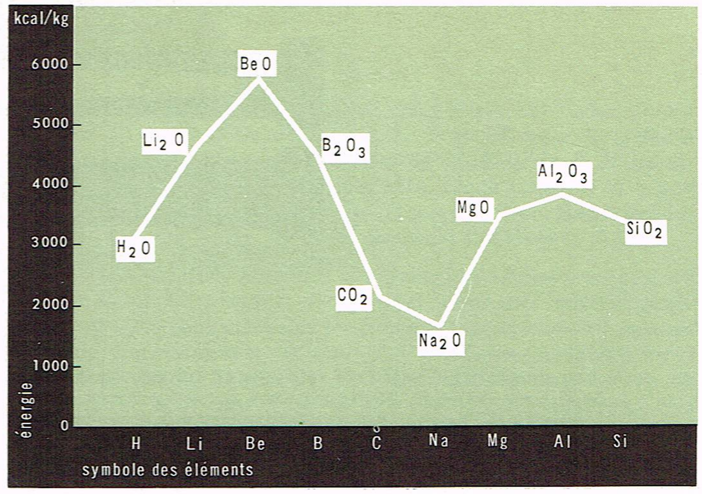
dans laquelle η est la fraction transformée en énergie cinétique ($\eta < 1$) et Q la quantité d'énergie libérée par unité de masse participant à la réaction.

Toutes choses égales, la vitesse d'éjection sera d'autant plus grande que la valeur de Q sera élevée, c'est-à-dire que l'on aura affaire à une source d'énergie plus puissante. La recherche des moyens permettant d'obtenir à l'aide de source d'énergie chimique des vitesses d'éjection élevées est donc liée à celle de réactions chimiques libérant, par unité de masse participant à la réaction, une quantité d'énergie aussi grande que possible. L'énergie libérée par une réaction chimique provient d'un réarrangement des atomes et des molécules par rapport à l'arrangement initial, c'est-à-dire celui des corps mis en présence dans la réaction.

On sait que seuls les électrons appartenant aux couches externes des atomes participant à la réaction jouent un rôle dans celle-ci. Il en résulte donc que, d'une façon générale du moins, ce seront les réactions faisant intervenir des éléments légers (ceux de la première ou à la rigueur de la seconde ligne de la classification de Mendeleiev) qui libéreront les plus grandes quantités d'énergie par unité de masse participant à la réaction.

H							He
Li	Be	B	C	N	O	F	Ne
Na	Mg	Al	Si	P	S	Cl	Ar

Les réactions les plus énergétiques sont celles qui interviennent entre les éléments de gauche du tableau (électropositifs et réducteurs) et ceux de droite (électronégatifs et oxydants).



Énergies libérées par des combustions avec l'oxygène.

Par exemple, l'oxydation de l'hydrogène ou du béryllium par l'oxygène ou le fluor.

$$H_2 + 1/2\ O_2 = H_2O$$
$$Be + O = BeO$$
$$F_2 + H_2 = 2\ FH.$$

Les quantités d'énergie libérées par unité de masse, par combinaison avec l'oxygène des premiers éléments, sont indiquées sur le graphique ci-dessus.

les fusées à propergols liquides

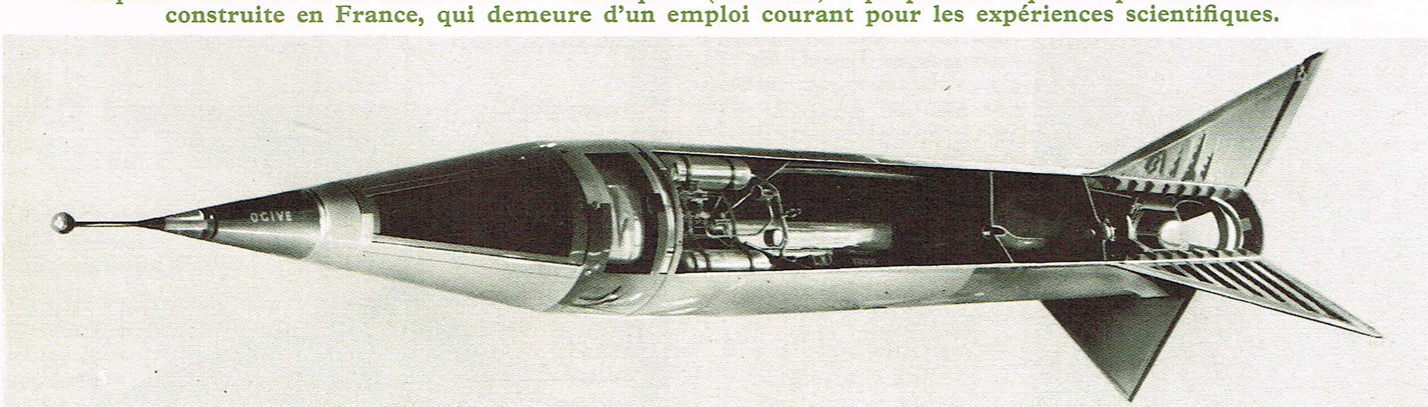
Les premières réalisations de fusées à liquides datent des années 1930 : c'est en 1938 exactement qu'eut lieu le premier vol de fusée allemande A3, qui n'était autre que la maquette du V2. La Seconde Guerre mondiale et surtout l'après-guerre ont vu se développer un peu partout dans le monde un nombre considérable de propulseurs de ce genre, de dimensions toujours plus imposantes et construits en vue de programmes toujours plus ambitieux. Voici une vingtaine d'années, le V2, dont le poids au départ était de l'ordre de 12 tonnes, laissait déjà rêveur le profane qui bien souvent entendait parler de fusée pour la première fois. On s'est accoutumé depuis à beaucoup plus gros : l'« Atlas », le « Titan », la fusée soviétique « T3A » par exemple; pourtant, leur poids au départ est de l'ordre de 100 tonnes pour les fusées américaines et supérieur à 200 tonnes pour les fusées soviétiques. Sans doute d'ici quelques années serons-nous aussi familiarisés avec les véhicules de la classe des projets américains « Nova », prévus pour des poids au départ de plusieurs milliers de tonnes, qui, lorsque nous y réfléchissons aujourd'hui, ne laissent pas néanmoins de nous impressionner.

un exploit peu banal

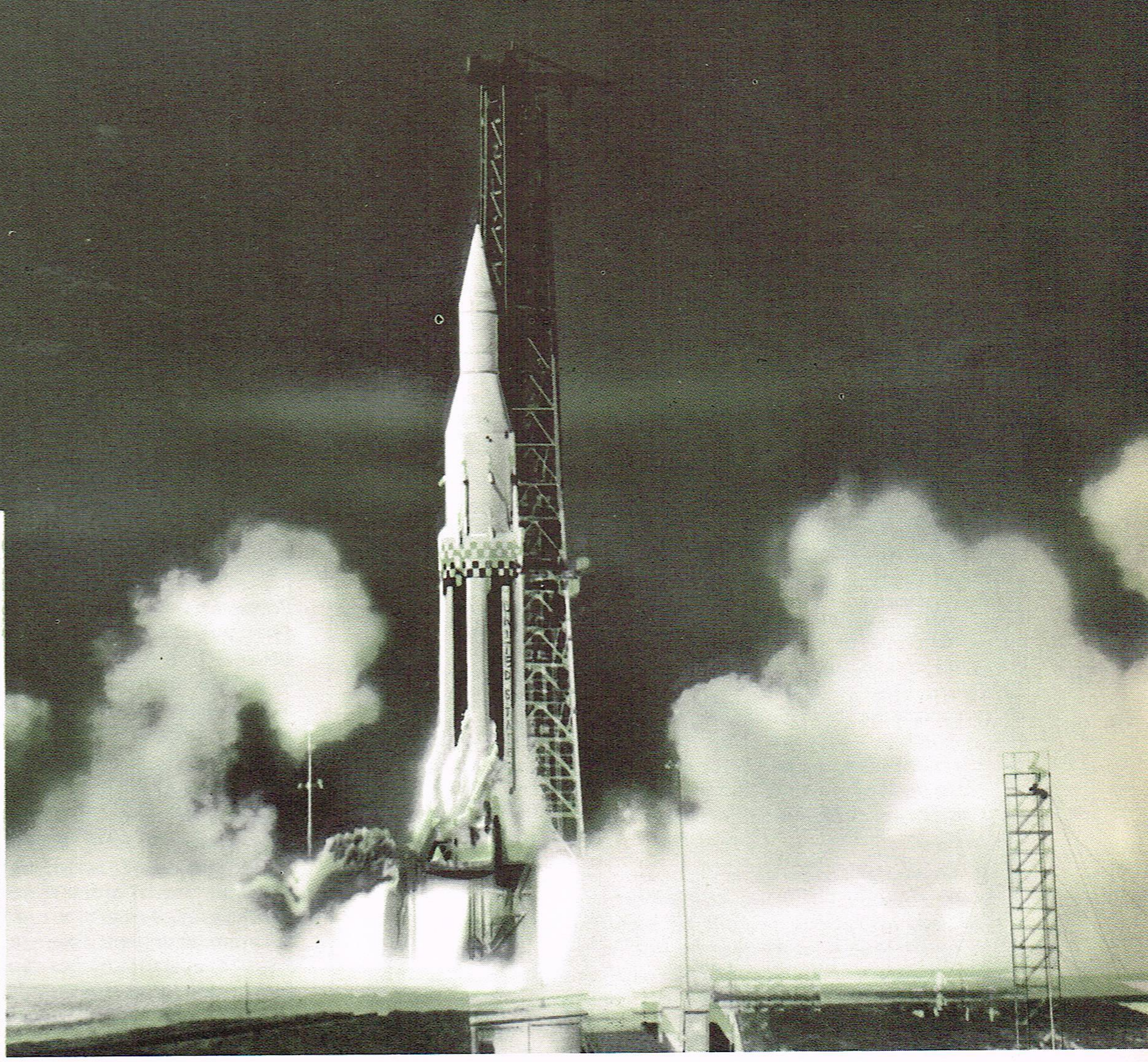
Prenons comme exemple le premier étage de « Saturn C1 », prévu entre autres pour les missions vers la Lune entreprises par les Américains (programme « Apollo »).

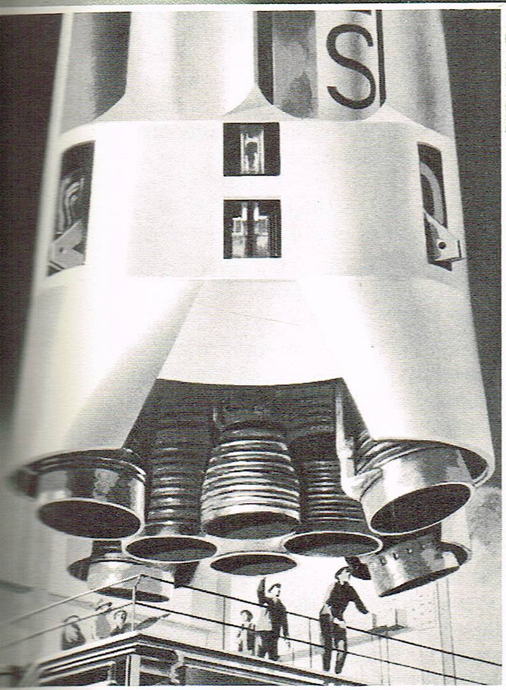
La vitesse d'éjection des gaz atteint 2 700 m/s et la poussée 700 tonnes. Toutes les secondes, la fusée consomme 1 500 litres d'oxygène et 1100 de kérosène, soit 2 600 litres en tout; la

Maquette écorchée de la fusée-sonde « Véronique » (L.R.B.A.) à propulsion liquide : première fusée-sonde construite en France, qui demeure d'un emploi courant pour les expériences scientifiques.



Phot. Larousse.





Lancement d'une fusée « Saturn ». *A gauche*, **ses huit tuyères.**

puissance nécessaire au seul fonctionnement des turbopompes alimentant les chambres de combustion est voisine de 30 000 kW, et supérieure à celle qui est disponible sur chacun des quatre arbres moteurs du paquebot *France;* le poids total de propergol transporté est d'environ 300 tonnes, ce qui, pour une densité moyenne voisine de l'unité, correspond à un volume total de réservoir de 300 000 litres. (Le moteur peut donc fonctionner pendant deux minutes environ.)

Intéressons-nous maintenant à ce qui se passe dans chacune des huit chambres de combustion : une puissance instantanée sous forme thermique, de l'ordre de 2 millions de kilowatts, y est développée, ce qui représente approximativement quatre fois la puissance thermique fournie par les plus grosses unités actuellement en service dans les centrales de l'E. D. F., mais ici elle est libérée dans une chaudière de quelques centaines de litres de volume et pesant une centaine de kilogrammes. Dans la deuxième version, actuellement en cours d'essais au banc, les huit chambres de combustion sont remplacées par une seule : la puissance libérée dans celle-ci est donc de 16 millions de kilowatts.

Lancer un « Saturn » représente donc l'exploit peu banal de communiquer à une gigantesque citerne de 300 m³ une vitesse de plusieurs milliers de mètres par seconde au moyen d'une source d'énergie développant une puissance comparable à la puissance totale installée dans l'ensemble des centrales thermiques françaises. Par ailleurs, bien qu'apparemment la fusée semble consister essentiellement en un immense réservoir, la réussite de son lancement exige le fonctionnement sans défaillance de plusieurs milliers d'éléments de tout ordre : mécanique, hydraulique, électrique, électronique et autres, miniaturisés et allégés au maximum, dont dépendent la propulsion, le guidage, le pilotage et le comportement de la charge utile.

L'art de l'ingénieur, dans ce domaine, consiste donc à apprendre à mettre en œuvre des sources d'énergie toujours plus puissantes, à réaliser des structures légères en mettant au point des techniques souvent absolument nouvelles, voire révolutionnaires, à manipuler, former et usiner des matériaux nouveaux, à construire des groupes de puissance pour un poids au kilowatt cent fois inférieur à celui qu'on réalise couramment dans l'industrie, à mettre en œuvre enfin les moyens nécessaires pour contrôler et piloter à chaque instant des véhicules dont les dimensions peuvent être gigantesques.

quel propergol employer ?

Les considérations développées plus haut en ce qui concerne les sources d'énergie l'ont déjà été il y a plusieurs dizaines d'années par Godard, Saenger, Esnault-Pelterie, et d'autres. Depuis cette époque et tout particulièrement depuis la fin de la dernière guerre, un très grand nombre de combinaisons ont été suggérées, étudiées et expérimentées. Mais celles qui ont fait l'objet d'applications pratiques pour des fusées réellement construites et livrées à l'utilisateur sont en nombre relativement

Phot. U. S. I. S.

restreint, **et il importe de préciser dès le départ que les performances les plus spectaculaires, réalisées aussi bien par les Russes que par les Américains, ne l'ont pas été à l'aide de « propergols miracles », mais en faisant brûler un produit pétrolier bon marché, le kérosène, avec de l'oxygène.**

D'une façon générale, les premiers constructeurs de fusées à liquides, probablement parce que leurs desseins n'étaient pas très ambitieux, ont davantage cherché à utiliser des produits industriels relativement peu coûteux et de manipulation facile, que des produits qui n'étaient et ne sont encore pour certains d'entre eux que des curiosités de laboratoire.

Il convient d'insister d'autre part sur le fait que tous les premiers étages à liquides utilisés jusqu'à présent comme lance-satellite sont des fusées militaires pour lesquelles on a davantage recherché la sécurité de fonctionnement et la facilité de mise en œuvre que des performances extraordinaires.

Il est probable que le développement des activités purement spatiales aura, et a déjà d'ailleurs, une conséquence dans la façon de concevoir une fusée et dans le choix des propergols. Eu égard aux difficultés inhérentes à la mise en œuvre des propergols à haute impulsion spécifique, il ne peut s'agir toutefois que d'une évolution lente, et pour de nombreuses années encore on préférera l'utilisation de propergols de performances peut-être modestes, mais de fonctionnement sûr (en particulier pour toutes les missions prévues avec des hommes à bord d'un astronef).

Il est facile de dresser une liste des propriétés souhaitables pour un combustible et un comburant, mais nous allons voir que la liste est longue et qu'il est bien difficile de trouver des corps ayant toutes les vertus. En fait, le choix d'un propergol résulte toujours d'un compromis entre ses qualités d'ordre énergétique et certaines autres qui conditionnent sa facilité d'utilisation.

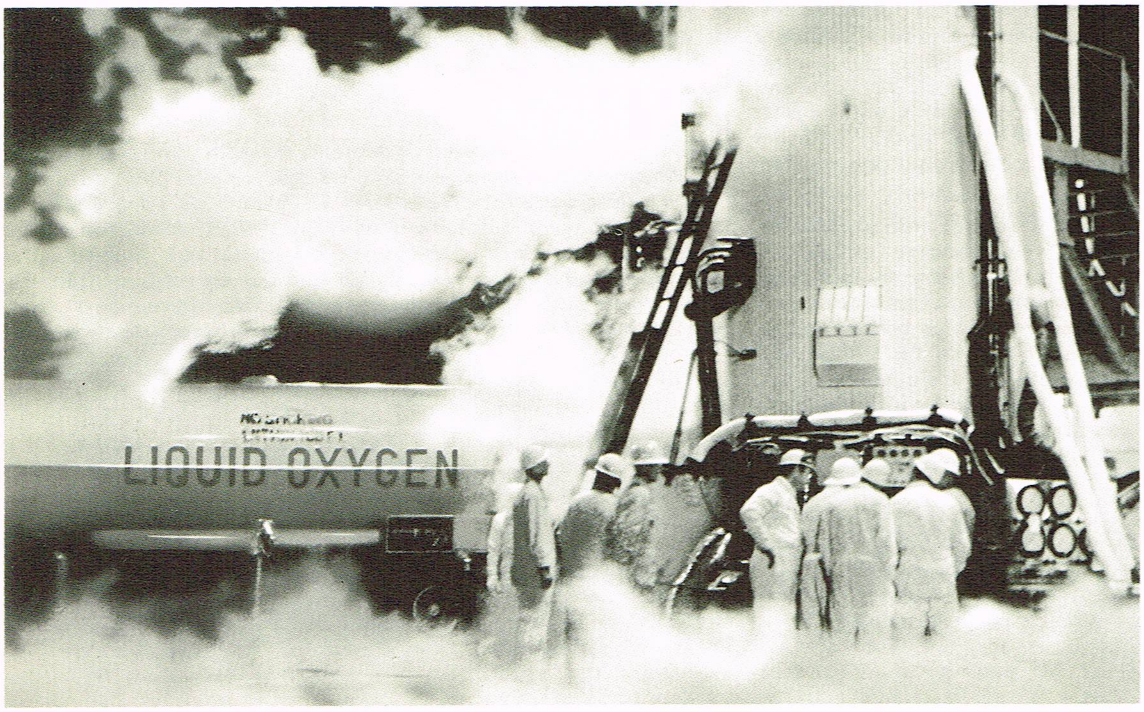
Par propriétés d'ordre énergétique, nous entendons celles dont dépendent les performances de la fusée, ce sont :
— l'impulsion spécifique de la combinaison utilisée;
— la densité moyenne du propergol (compte tenu des masses respectivement en présence de l'un et l'autre composant).

Les propriétés qui en conditionnent la mise en œuvre sont :
— la température d'ébullition,
— la stabilité,
— le comportement à l'égard des matériaux,
— la toxicité, et les dangers présentés par la manipulation des produits,
— le prix.

A peu d'exceptions près, les propergols possédant les meilleures propriétés énergétiques sont ceux dont la mise en œuvre est la plus difficile : le fluor et les composés fluorés, par exemple, forment en général avec les mêmes combustibles des combinaisons plus énergétiques que l'oxygène et les composés oxygénés. Malheureusement, le fluor et ses porteurs sont tous des agents extrêmement corrosifs, de manipulation très dangereuse, et tous les efforts consacrés jusqu'ici à l'utilisation de produits fluorés n'ont pas encore permis la mise au point d'un seul moteur fonctionnant dans des conditions satisfaisantes.

Comme carburant on a essentiellement utilisé jusqu'ici l'oxygène pur (sous forme liquide) ou des porteurs d'oxygène tels que :
— l'acide nitrique NO_3H,
— le peroxyde d'azote N_2O_4,
— l'eau oxygénée concentrée H_2O_2.

L'oxygène liquide, comme le fluor liquide d'ailleurs, présente l'inconvénient d'avoir un point d'ébullition extrêmement bas (respectivement — 183 et — 188 °C). Dans un réservoir peu isolé, voire non isolé thermiquement, comme c'est généralement le cas pour une fusée, un échange de chaleur considérable s'effectue avec l'atmosphère; ce qui conduit à un taux d'évaporation important dans le réservoir. Cet effet n'est certes pas suffisant pour entraîner la perte d'une quantité appréciable d'oxygène pendant les quelques minutes de fonctionnement de la fusée; il exclut toutefois la possiblité de conserver celle-ci prête au départ sur son aire de lancement. Pour les missions d'astronautique auxquelles nous nous intéressons ici, la sujétion d'avoir à remplir le réservoir d'oxygène seulement dans l'heure ou dans les heures qui précèdent le départ ne présente pas d'inconvénient grave; elle en présente, par contre, pour des missions militaires exigeant la possibilité d'une attaque et surtout d'une riposte instantanée. **C'est la raison pour**



Remplissage du réservoir d'oxygène liquide d'une fusée.

Phot. U. S. I. S.

laquelle les engins balistiques « Atlas » (à oxygène liquide) sont progressivement remplacés par des engins pouvant être conservés indéfiniment prêts au départ, tels que le « Minuteman » (à propergol solide) ou le « Titan » (à propergol liquide stockable).

En ce qui concerne l'effet corrosif, la toxicité, et d'une façon générale la mise en œuvre, les porteurs d'oxygène présentent incomparablement moins de difficultés que les produits fluorés : des produits tels que l'acide nitrique pur ou l'eau oxygénée concentrée à 95 p. 100 sont aujourd'hui manipulés couramment et font, de ce fait, l'objet de nombreuses applications.

Ce sont également surtout des combustibles stockables et de mise en œuvre facile, tels que les produits pétroliers et d'autres produits organiques, qui ont été principalement utilisés comme combustibles. L'hydrogène pur peut, certes, donner, avec les mêmes comburants, des impulsions spécifiques beaucoup plus élevées que celles qui sont obtenues en l'utilisant par l'intermédiaire de produits organiques tels que les carbures saturés. Il présente toutefois deux inconvénients importants :
— son point d'ébullition encore plus bas que celui de l'oxygène (— 253 °C);
— sa densité extrêmement faible (0,07 sous forme liquide), qui limite ses possibilités d'utilisation à bord d'un premier étage, l'économie de poids par rapport à l'emploi d'un propergol d'impulsion spécifique plus élevée étant obtenue au prix d'une augmentation non négligeable des dimensions du premier étage, de son poids à sec et donc de son prix.

Par contre, son emploi à bord d'étages supérieurs (toujours de dimensions plus modestes que les premiers étages) présente moins d'inconvénients à ce point de vue et permet d'augmenter considérablement la charge utile transportée. Une partie importante du programme spatial américain est fondée sur l'emploi de lanceurs dont les étages supérieurs utilisent l'hydrogène comme combustible. Ce sont « Atlas-Centaur » et les véhicules de la famille « Saturn », ces derniers étant plus particulièrement étudiés en vue de la réalisation du programme « Apollo », qui a pour but l'envoi d'hommes sur la Lune. Plusieurs lancements d' « Atlas-Centaur » et de « Saturn » ont déjà été effectués avec succès. La même tendance vers l'emploi de l'hydrogène dans les étages supérieurs se retrouve d'ailleurs dans les programmes spatiaux en dehors des Etats-Unis.

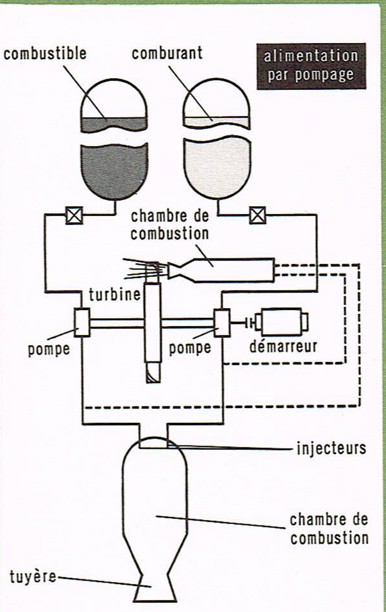
quelques problèmes de construction et de réalisation

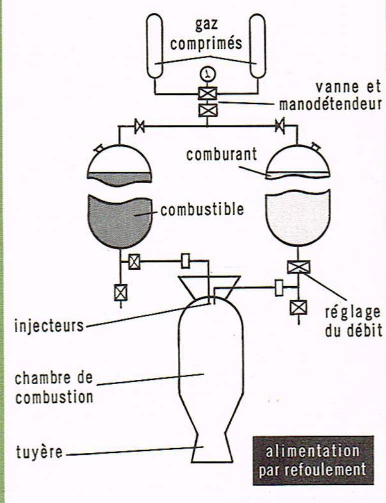
Ils sont évidemment nombreux et de tous ordres, mais tous ont en commun de devoir être résolus en recherchant simultanément l'allégement maximal et la meilleure sûreté de fonctionnement. Une discussion approfondie sur ce sujet relève évidemment d'un ouvrage spécialisé, aussi nous limiterons-nous à évoquer simplement ces problèmes.

alimentation de la chambre de combustion

La pression dans la chambre de combustion est généralement élevée, de l'ordre de la dizaine ou de quelques dizaines

L'alimentation par une pompe est prévue pour toutes les fusées de dimensions importantes. Nous avons vu plus haut que les puissances mises en jeu peuvent être considérables : plusieurs milliers, voire plusieurs dizaines de milliers de kilowatts. Les pompes sont entraînées par une turbine alimentée par un générateur de gaz chauds utilisant généralement les mêmes propergols que ceux qui sont utilisés pour la propulsion.





L'alimentation du générateur est elle-même effectuée par les pompes principales. Le démarrage de l'ensemble suppose évidemment l'existence d'un moyen quelconque permettant de mettre la turbine en marche au moins une fois : cartouche de poudre, démarreur électrique, pneumatique ou autre.

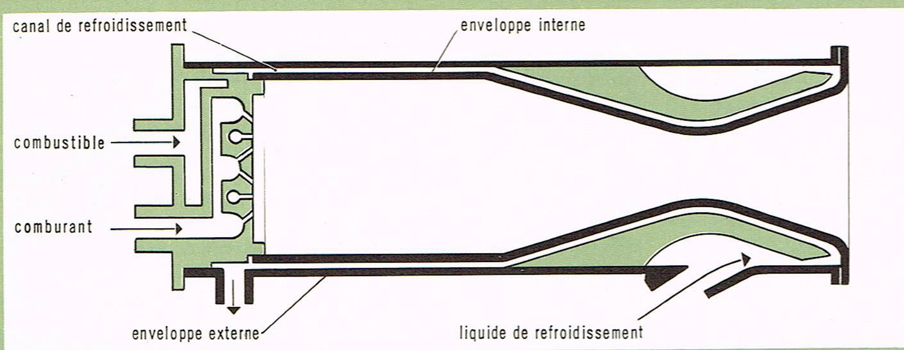
LES PRINCIPAUX PROPERGOLS UTILISÉS

Combustibles \ Comburants	O_2 liquide	N_2O_4	NO_3H	H_2O_2	F_2 liquide	ClF_3
Hydrogène liquide	2 985 391	2 785 341	2 540 326	2 300 314	3 960 410	3 420 318
Porteurs d'hydrogène						
1. Hydrazine et dérivés						
Hydrazine N_2H_4	3 400 313	3 210 292	3 045 283	2 840 282	4 660 364	3 905 292
Monométhylhydrazine $N_2H_3(CH_3)$	3 580 312	3 390 288	3 165 279	2 910 279	4 390 346	3 710 284
Diméthylhydrazine $N_2H_2(CH_3)_2$	3 615 310	3 435 286	3 230 277	2 925 278	4 290 243	3 660 278
2. Produits pétroliers						
Kérosène avion	3 670 301	3 450 276	3 235 268	2 910 272	4 410 326	3 505 258
3. Alcools						
Alcool éthylique $C_2H_5(OH)$	3 423 288	3 236 269	2 940 250		4 283 336	
Alcool méthylique $CH_3(OH)$			2 830 249			

Pression de combustion : 68 atm.

Pression à la sortie de la tuyère : 1 atm.

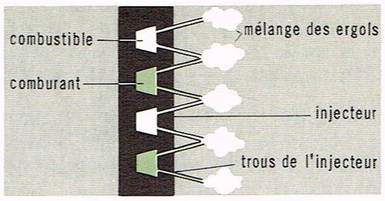
La première ligne indique (en degrés Kelvin) la température de combustion; la seconde, l'impulsion spécifique.

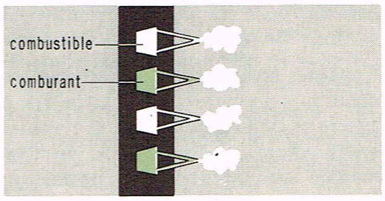


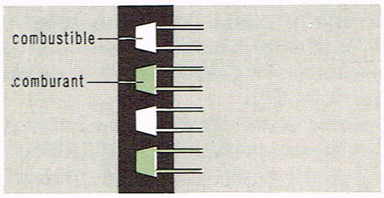
Refroidissement par circulation dans une double enveloppe.

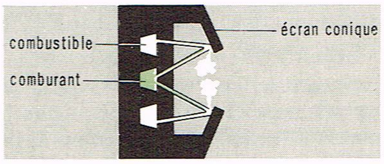
Les matériaux généralement utilisés pour la construction des tuyères sont : les alliages légers, l'acier inoxydable, le nickel. Pour des fusées de petites dimensions et en vue d'applications assez particulières, on a expérimenté des tuyères en tungstène ou en d'autres métaux ou alliages à haut point de fusion ne nécessitant aucun système de refroidissement. Eu égard aux densités du tungstène ou des matériaux de ce genre, une telle solution conduit évidemment à des poids rédhibitoires, sauf pour des fusées de très faible poussée. Pour les matériaux cités plus haut, il est nécessaire de prévoir un système de refroidissement. C'est toujours l'un des deux liquides qui est utilisé à cette fin. Dans les fusées de faible ou moyenne poussée, on adopte généralement la solution figurée ci-dessus, consistant à faire circuler le liquide à l'intérieur d'une double enveloppe. Dans le cas de fusées à très forte poussée («Thor», «Atlas», etc.) l'ensemble de la tuyère et de la chambre de combustion est constitué d'un faisceau de tubes en nickel extrêmement fins (de l'ordre de quelques dixièmes de millimètre d'épaisseur), le liquide de refroidissement se répartissant également dans les tubes. Chaque tube, circulaire au départ, est déformé de façon telle qu'il y ait contact entre tous les tubes tout le long du profil de la tuyère. L'ensemble des tubes est brasé et contenu à l'intérieur d'anneaux dont le rôle est de supporter les efforts mécaniques engendrés par les pressions élevées régnant à l'intérieur du moteur.

Quelques configurations d'injecteurs.









« Pomme d'arrosoir » portant les nombreux trous par où se fait l'injection des deux ergols.



d'atmosphères. Pour y injecter les propergols, il est donc nécessaire de disposer d'un moyen permettant de mettre ceux-ci sous une pression supérieure à celle qui règne dans la chambre.

Cette mise en pression peut s'effectuer de deux façons :

— à l'aide de pompes;

— en mettant les réservoirs de propergols eux-mêmes sous pression.

échanges thermiques dans la chambre de combustion et la tuyère

Les températures de combustion des propergols généralement utilisés sont au moins de l'ordre de 3 000 °K, voire supérieures.

Il en résulte des échanges thermiques élevés, et dans tous les cas la nécessité de prévoir un moyen pour maintenir les matériaux constituant les parois de la tuyère et de la chambre de combustion à une température compatible avec leur bonne tenue mécanique.

dispositif d'injection

De la qualité du dispositif d'injection dépendent la stabilité et aussi la qualité de la combustion dans la chambre, le but étant d'obtenir une impulsion spécifique réelle aussi proche que possible de celle que l'on calcule dans les conditions idéales (valeurs données dans le tableau précédent). En général, l'injection se fait par un grand nombre de trous de faible diamètre (douche ou pomme d'arrosoir) disposés sur des cercles concentriques en alternant les trous comburant et les trous combustible. De nombreuses autres configurations peuvent toutefois être envisagées.

pilotage

On ne peut évidemment pas bénéficier d'un effet aérodynamique pour stabiliser et pour piloter un véhicule évoluant hors de l'atmosphère : il est nécessaire de prévoir d'autres moyens. La stabilisation en roulis (mouvement autour de l'axe du véhicule) s'effectue, dans la plupart des cas, par l'intermédiaire de petites fusées annexes et son pilotage (c'est-à-dire le contrôle de sa trajectoire) par déviation du jet du propulseur principal.

Phot. X.

les fusées à propergols solides

Substituer aux deux réservoirs de liquides, un bloc de poudre de composition, de forme et de dimensions choisies, supprimer ainsi le dispositif d'alimentation et toutes les sujétions liées à l'emploi de liquides de manipulation difficile est sans doute une solution séduisante. De ce point de vue, le retour à la conception première des fusées représente effectivement un progrès.

Toutefois, les qualités que doit réunir un tel mélange sont nombreuses :
— il doit être stable dans les conditions de stockage habituelles, et ne pas se transformer dans le temps ou sous l'influence de variations de température par exemple;
— la combustion ne doit pas s'effectuer de façon désordonnée et aléatoire, mais, au contraire, avec la vitesse de combustion souhaitée, et en conservant au bloc une forme géométrique telle que la quantité de gaz libéré par seconde soit aussi constante que possible;
— enfin, bien entendu, il est souhaitable que la quantité d'énergie libérée par unité de masse de poudre soit aussi élevée que possible.

Les corps susceptibles d'entrer dans la composition d'une poudre ne permettent malheureusement pas d'obtenir le mélange idéal, et la réalisation d'un tel propergol est nécessairement le fruit d'un compromis entre les qualités d'ordre énergétique que l'on désire obtenir et certaines sujétions limitant le choix de matériaux susceptibles d'entrer dans la composition de la poudre. Il en résulte que, d'une façon générale, les propergols solides ont de moins bonnes performances que les propergols liquides, mais que, par contre, ils conduisent à des fusées de conception plus simple et de mise en œuvre plus facile. Pour ces raisons, ils ont été, au moins aux Etats-Unis, pour une large part préférés aux propergols liquides pour les missions militaires : les « Polaris » transportés par les sous-marins atomiques américains sont des fusées de moyenne portée à poudre, ainsi que les « Minuteman » (longue portée) et les « Pershing » (courte portée).

poudres colloïdales et poudres composites

Dans les conditions ordinaires de température et de pression, une poudre doit être chimiquement stable malgré le mélange des différents corps qui la constituent, et la réaction de décomposition ne doit pouvoir être amorcée que dans certaines conditions de pression et de température. Cette réaction est appelée « combustion de la poudre ». Elle doit pouvoir s'entretenir d'elle-même, être continue et reproductible. Elle se propage perpendiculairement à la surface de la poudre à une vitesse faible : quelques millimètres, au plus quelques centimètres, par seconde. L'usage a prévalu de classer les principaux propergols solides en deux grandes familles : les poudres colloïdales déjà relativement anciennes et les poudres composites auxquelles se rattachent à la fois les poudres très anciennes, telle la poudre noire, et les poudres les plus évoluées.

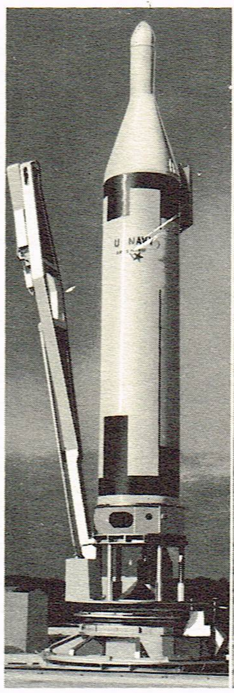
On obtient le plus souvent les poudres colloïdales en incorporant dans un support nitré, généralement la nitrocellulose, une seconde substance explosive, généralement la nitroglycérine, et en gélatinisant le tout.

Dans les poudres composites, au contraire, l'oxydant est introduit sous forme cristallisée, finement broyé; les oxydants classiques sont les nitrates ou les perchlorates d'ammonium ou de métaux alcalins.

La poudre noire a été longtemps le seul propergol connu. Elle n'est plus utilisée que pour l'allumage, car les blocs de poudre noire sont fragiles, leur combustion donne une forte proportion de résidus solides et l'impulsion spécifique correspondante est faible.

Dans les poudres composites modernes, le combustible forme liant. Celui-ci est, en général, une résine synthétique moulable et thermodurcissable : par exemple, le polyuréthane ou d'autres produits du même genre. Ces poudres se prêtent assez bien au moulage de blocs de toutes dimensions.

Les impulsions spécifiques qui caractérisent ces poudres, même les plus évoluées, sont sensiblement inférieures à celles qui peuvent être obtenues avec les propergols liquides, même de performances modestes, tels que les propergols stockables couramment utilisés aujourd'hui. Cela est dû à la proportion importante d'atomes d'éléments lourds (Cl) ou inertes (N) et à la concentration relativement faible en hydrogène. On améliore considérablement les performances en diminuant autant que possible la quantité de liant et en substituant à la proportion éliminée des éléments légers sous forme pulvérulente tels que le béryllium, le lithium ou l'aluminium. Les poudres actuellement



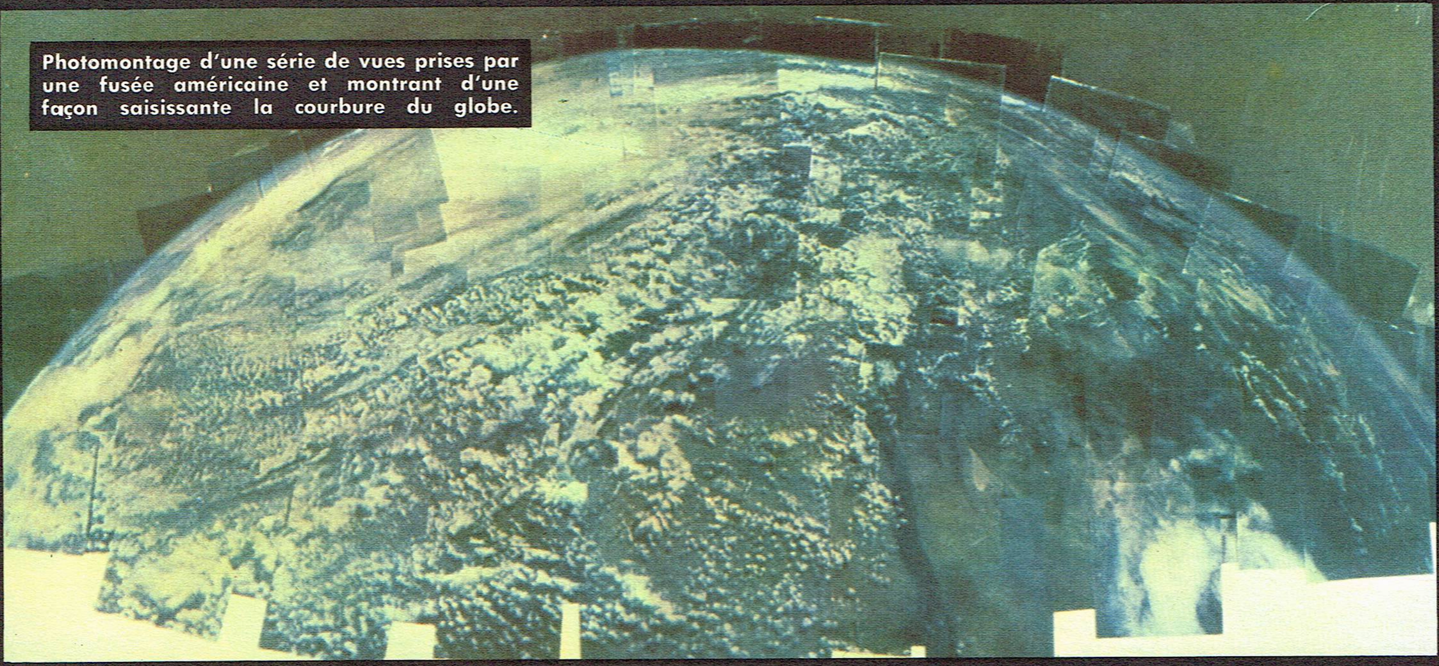


Fusée « Polaris » au sol et lancée par un sous-marin en plongée.

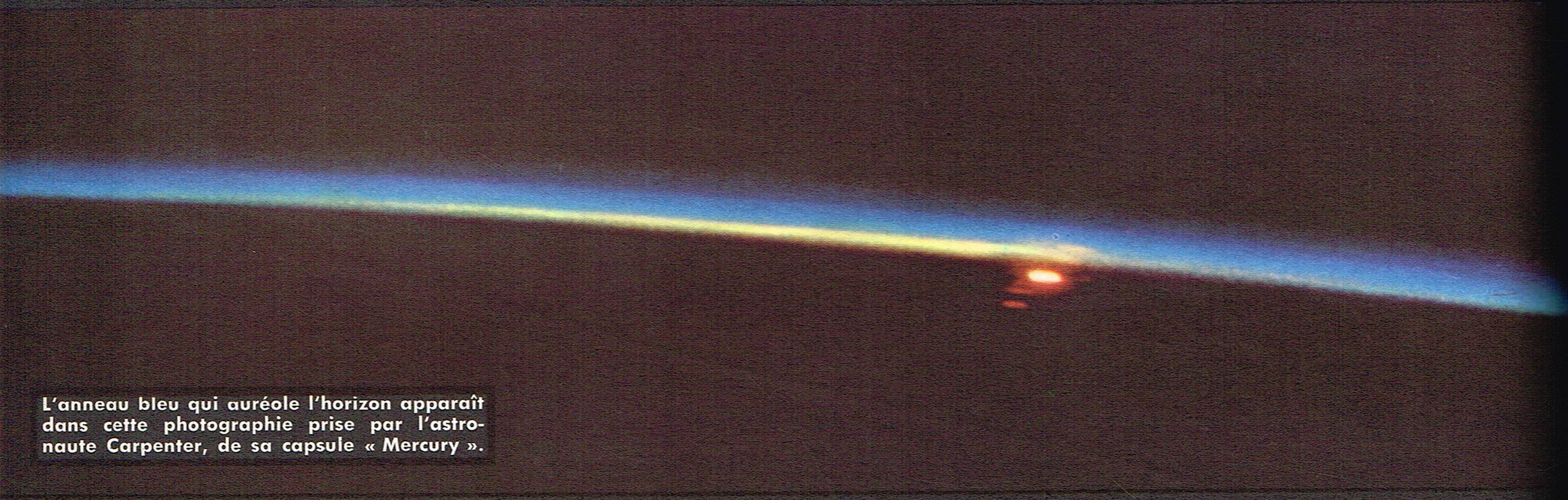
COMPOSITION DE QUELQUES POUDRES

poudre noire	**salpêtre**	**40 p. 100**
composition indiquée par le moine Roger Bacon (1260)	**soufre**	**30 p. 100**
	charbon fin de noisetier	**30 p. 100**
composition actuelle	**salpêtre**	**70 p. 100**
	charbon	**10 à 20 p. 100**
	soufre	**10 à 20 p. 100**
poudre colloïdale	**nitrocellulose**	**66 p. 100**
utilisée dans l'armement	**nitroglycérine**	**25 p. 100**
	centralite	**8 p. 100**
	divers	**1 p. 100**
poudre composite :		
	perchlorate d'ammonium	**80 p. 100**
exemple de composition	**résine synthétique**	**20 p. 100**
poudre composite chargée :		
	perchlorate d'ammonium	**66 p. 100**
exemple de composition	**polyuréthane**	**17 p. 100**
	aluminium (pulvérulent)	**17 p. 100**

Phot. U. S. I. S.



Photomontage d'une série de vues prises par une fusée américaine et montrant d'une façon saisissante la courbure du globe.



L'anneau bleu qui auréole l'horizon apparaît dans cette photographie prise par l'astronaute Carpenter, de sa capsule « Mercury ».



La côte nord-ouest de l'Afrique photographiée par Carpenter.

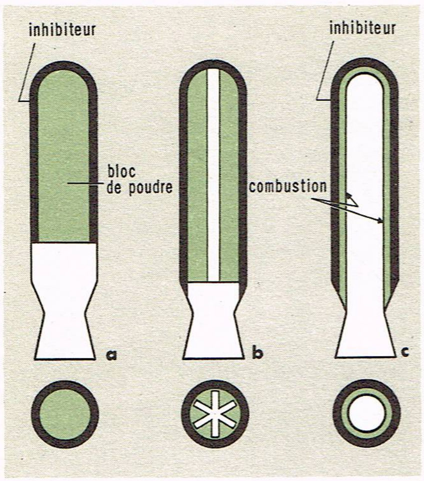


Premier lancement expérimental de la fusée « Saturn ».



Départ de la fusée « Thor-Able », qui mettra sur orbite le premier satellite « Tiros ».

Phot. U.S.I.S.



Blocs de poudre :
a) combustion frontale et poussée constante ;
b) combustion radiale et poussée constante ;
c) combustion radiale et poussée progressive.

Schéma d'un allumeur.

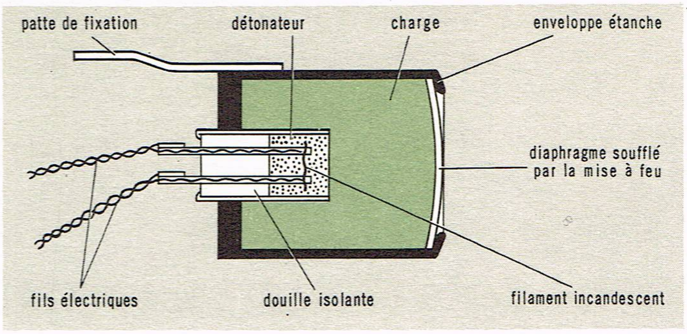
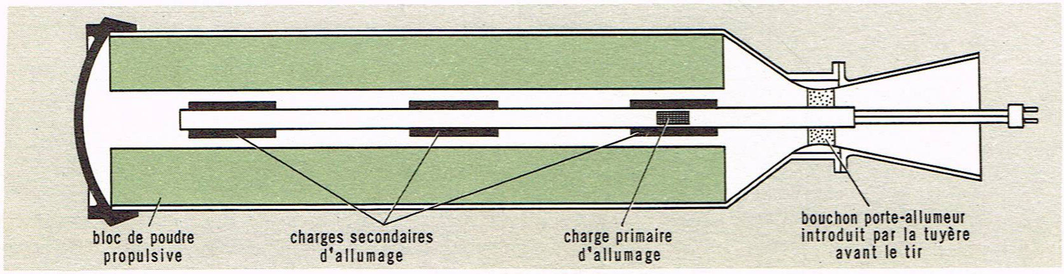


Schéma type d'un dispositif d'allumage.



Une masse de poudre d'allumage est répartie en une ou plusieurs charges convenablement disposées à l'intérieur du propulseur. Une seule de ces charges est munie d'inflammateurs constitués essentiellement par une résistance électrique. Le passage du courant provoque la montée quasi instantanée de la température de chacune des charges, de leurs inflammateurs et, par voie de conséquence, celle du bloc.

utilisées pour la propulsion des fusées du type « Polaris » ont une composition de ce genre dont les proportions sont données en quatrième colonne du tableau précédent.

dessin d'un bloc de poudre

Dans le cas d'une fusée à poudre, la quantité q qui intervient dans l'expression de la poussée $F = q.V$ est évidemment égale à la quantité de poudre brûlée par unité de temps. De la même façon que l'on peut définir la vitesse à laquelle une cigarette se consume dans l'air, on peut définir une vitesse de combustion de la poudre que nous appellerons v. Considérons, par exemple, un bloc de poudre à surface de combustion frontale brûlant précisément comme une cigarette :

A l'instant initial, la surface du bloc est dans le plan transversal passant par A; au bout d'un certain temps t, elle est dans le plan passant par A' : $v = AA'/t$.

Si S est la surface de combustion, $q = Sv.\rho$ (ρ étant la masse spécifique de la poudre) et l'expression de la poussée devient $F = \rho.S.v.V$.

La poussée d'une fusée est donc proportionnelle à la surface instantanée de combustion et à la vitesse de propagation de celle-ci. Pour une pression donnée, cette dernière ne dépend pratiquement que de la nature de la poudre; elle est en général de l'ordre de quelques millimètres par seconde.

La poussée provenant de la combustion d'un bloc de ce genre est constante puisque la surface S de combustion ne varie pas. Par contre, il n'est pas possible d'obtenir de cette façon des poussées importantes. Supposons, par exemple, que l'on utilise une poudre de masse spécifique 1 kg/dm^3, de vitesse de combustion 10 mm/s, et avec un rapport de détente tel que la vitesse d'éjection soit 2 500 m/s. La poussée sera égale à

$$F = S \times 0{,}01 \times \frac{1000}{9{\cdot}81} \times 2\,500 \text{ (S étant exprimé en m}^2\text{)}.$$

A une poussée de 2 500 kg correspondra donc approximativement une surface frontale de 1 m^2; pour une fusée de 100 tonnes de poussée, il faudrait 40 m^2 de surface frontale, ce qui est pratiquement irréalisable.

Pour augmenter la surface de combustion, tout en conservant un diamètre de bloc raisonnable, on perfore les blocs dans le sens de la longueur et on utilise comme surface de combustion la surface intérieure ainsi définie. La forme de la perforation est choisie de façon telle que la surface de combustion demeure constante ou tout au moins respecte une certaine loi que l'on s'est fixée. On obtient ainsi des formes complexes comme celles qui sont représentées sur la figure ci-dessus.

l'inhibiteur

Considérons, par exemple, un bloc perforé à combustion interne. Si aucune disposition n'est prise pour empêcher la combustion par les faces latérales, celle-ci s'étendra rapidement à toute la surface du bloc. Outre le fait que ce mode de combustion ne correspondra pas à la loi de poussée désirée, il en résultera très rapidement une augmentation de la pression dans le propulseur et l'explosion de celui-ci. Pour limiter la surface de combustion, on revêt les faces par lesquelles celle-ci ne doit pas se produire d'une résine non ou peu combustible qui adhère solidement au bloc et que l'on appelle l' « inhibiteur ».

l'allumage

L'allumage d'une fusée à poudre est obtenu à l'aide de dispositifs spéciaux, dont la conception demeure souvent assez empirique. Ce dispositif, ou « allumeur », a pour but de faire l'apport initial d'énergie qui, en élevant la température de la poudre, provoquera l'inflammation de celle-ci. Un dispositif type est représenté sur la figure ci-dessus.

Jusqu'ici les blocs de poudre étaient fabriqués par des procédés mécaniques; il en résultait une limitation dans les dimensions, correspondant aux machines dont on pouvait disposer. Une nouvelle technique s'est développée depuis quelques années,

consistant à remplir le corps du propulseur directement avec la poudre sous forme pâteuse et à laisser celle-ci polymériser; aucune limitation de dimensions n'intervient si l'on opère de cette façon et des fusées à poudre de plusieurs mètres de diamètre sont actuellement en cours de réalisation ou d'essais.

quelques problèmes propres à la mise en œuvre des fusées à poudre

Indépendamment des problèmes liés à l'élaboration et à la mise en œuvre des propergols solides eux-mêmes, la construction d'un propulseur présente de nombreuses difficultés qui doivent toujours être résolues dans l'optique du poids minimal et de la sûreté de fonctionnement maximale. Parmi ceux-ci, on peut citer : la recherche de matériaux propres à être utilisés pour la construction de l'enceinte dans laquelle est enfermé le bloc de poudre, ainsi que celle des matériaux propres à être utilisés dans la construction de la tuyère, pour le refroidissement de laquelle on ne dispose généralement pas de liquide.

Dans le premier cas, on recherchera essentiellement à utiliser des matériaux de haute tenue mécanique, pour un poids aussi faible que possible. Les recherches effectuées dans ce dessein ont permis d'ailleurs de mettre au point des aciers ayant une charge de rupture extrêmement élevée, de l'ordre de 250 à 300 kg/mm², ou d'autres techniques absolument nouvelles, telles que l'emploi de fils de verre bobinés et imprégnés de résine.

Aux matériaux de la tuyère, on demande d'être capables de résister à une température de combustion supérieure à 3 000 °C et, le cas échéant, à l'érosion mécanique due aux particules condensées d'alumine provenant de la combustion pendant une durée pouvant être de l'ordre de la minute, voire plus élevée. Dans la région du col, où les échanges thermiques sont les plus importants, on emploie généralement des graphites spéciaux ou du tungstène.

les fusées à énergie nucléaire

Dans ce type de fusées, la source d'énergie est constituée par un réacteur nucléaire à fission, et la source de masse par une réserve d'hydrogène chauffé par passage au travers du réacteur nucléaire, puis détendu dans une tuyère.

Référons-nous à l'expression qui lie l'énergie dégagée dans une réaction chimique à la vitesse d'éjection des gaz provenant de celle-ci : $1/2\ m\ V_e^2 = \eta\ mQ$.

Il est clair que la quantité d'énergie libérée par unité de masse participant à la réaction est limitée lorsqu'elle est fournie par une réaction chimique. Par contre, elle peut être aussi grande qu'on le souhaite, au moins théoriquement, si elle est fournie par une source d'énergie extérieure. Pour un gaz donné, augmenter Q consiste, en fait, à augmenter sa température. Dans le cas d'une fusée à énergie nucléaire, la valeur Q sera donc limitée par des impératifs d'ordre technologique, entre autres par la tenue mécanique du réacteur, qui devra fonctionner à une température nécessairement plus élevée que celle du gaz chauffé pour céder de l'énergie à celui-ci, **mais non de façon intrinsèque comme dans le cas d'une réaction chimique.**

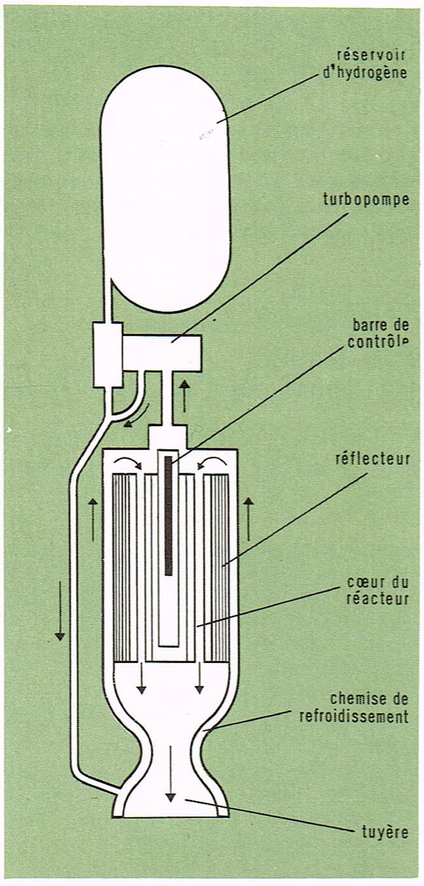


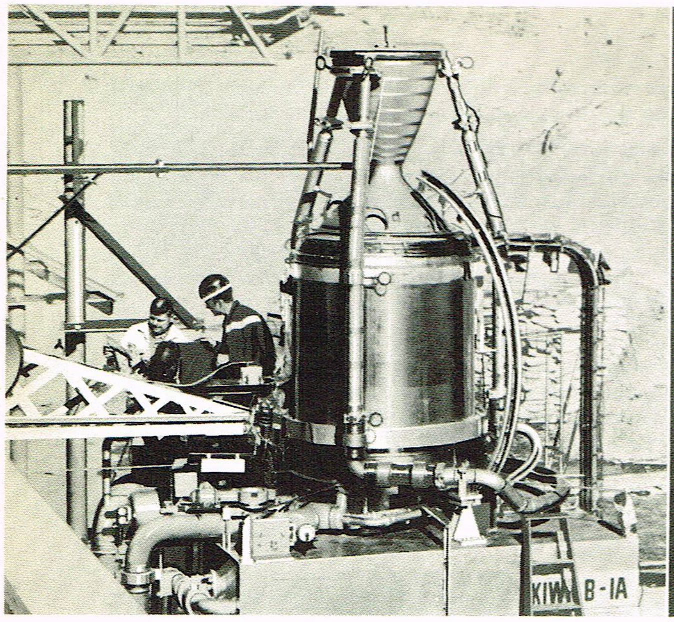
Schéma de principe d'un propulseur nucléaire.

Pour un même débit d'énergie fourni par le réacteur, la vitesse d'éjection sera d'autant plus élevée que la capacité calorifique du fluide moteur sera grande. Si T est la température maximale imposée par la bonne tenue mécanique du réacteur, l'énergie qui pourra être emmagasinée dans le fluide moteur à cette température et qui sera susceptible d'être transformée en énergie cinétique sera effectivement d'autant plus grande que la capacité calorifique de ce gaz (quantité d'énergie nécessaire à élever la température de l'unité de masse du gaz considéré de 1 degré) sera grande : on sait que la capacité calorifique d'un gaz est liée à sa structure moléculaire et non à la masse atomique des éléments qui constituent celle-ci. Ce sont donc de nouveau les corps constitués d'éléments légers — en particulier l'hydrogène — qui disposent d'une capacité calorifique élevée et dont l'emploi est avantageux en propulsion nucléaire.

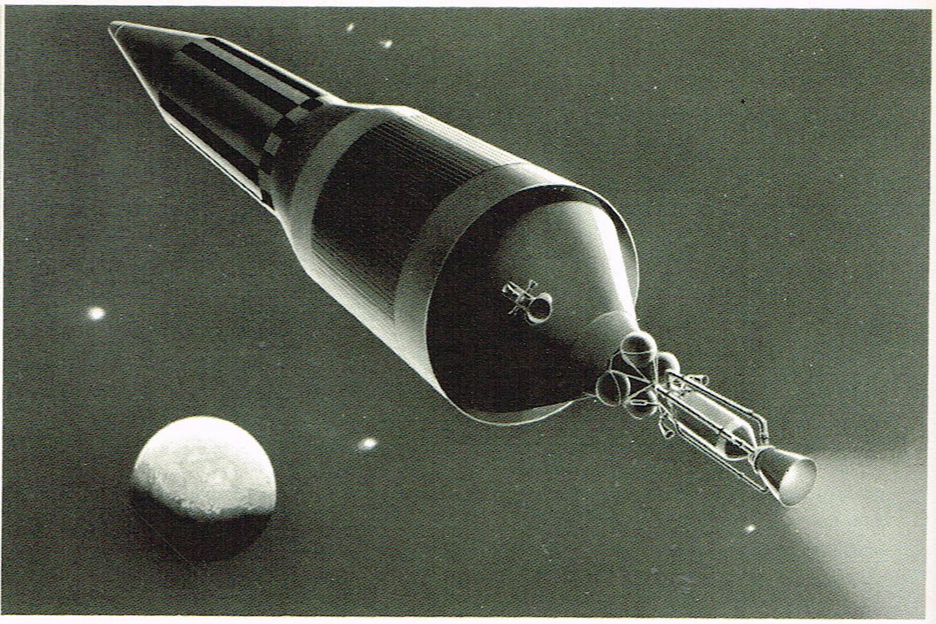
L'étude thermodynamique d'une telle fusée montre que, en utilisant l'hydrogène comme fluide moteur, il est théoriquement possible d'obtenir des impulsions spécifiques nettement plus élevées que celles qui sont réalisables avec des propergols chimiques (dans les mêmes conditions de détente, de l'ordre de 800 s à 1 000 s, avec une température de réacteur encore acceptable, c'est-à-dire de l'ordre de 2 500 à 3 000 °K).

Une fusée à énergie nucléaire est essentiellement constituée :
— d'un réservoir d'hydrogène;
— d'un dispositif d'alimentation permettant de fournir à l'hydrogène la pression suffisante pour qu'il puisse traverser les réacteurs et se détendre ensuite;
— d'un réacteur;
— d'une tuyère.

Un écran de protection peut être prévu à l'arrière du réacteur dans le but de protéger, le cas échéant, la charge utile contre le rayonnement des neutrons ou des rayons gamma en provenance de celui-ci.



Essais de réacteurs « Kiwi », destinés à la mise au point d'un propulseur pour fusées utilisant l'énergie nucléaire.



Dessin montrant l'aspect de l'engin « Nerva » à propulsion nucléaire.

le réacteur

Sa réalisation est évidemment le problème clé.

Pour une poussée de 100 tonnes et une vitesse d'éjection de 10 000 m/s, l'énergie cinétique contenue dans le jet correspond à une puissance de 7 millions de kilowatts. Compte tenu des différents rendements et des pertes dans le jet, la réaction doit fournir environ 5 millions de kilowatts, c'est-à-dire une puissance de l'ordre de grandeur de celle qui est fournie actuellement par les plus grosses centrales nucléaires.

Pratiquement, la propulsion ne présente donc d'intérêt que dans la mesure où l'on est capable de réaliser des réacteurs légers. On a calculé que pour concurrencer valablement les fusées à liquide il était nécessaire que le poids spécifique des réacteurs soit inférieur à 2 kg/MW et la puissance volumique au moins de plusieurs MW/litre. Or, nous verrons que la puissance volumique des réacteurs « thermiques » actuellement en service ou en projet, pour la fourniture d'énergie au sol, s'élève à quelques dizaines de KW/litre, celle des réacteurs à « neutrons rapides » étant de quelques centaines.

On sait que, au moins dans les cas des réacteurs de conception classique, ceux-ci sont essentiellement constitués d'un cœur comprenant le matériau fissile et le modérateur, au sein duquel circule le fluide dont le rôle est de prélever l'énergie libérée par la réaction de fission. Dans un réacteur de centrale, ce sera, par exemple, le gaz carbonique; dans le cas d'une fusée nucléaire, ce sera directement l'hydrogène, qui entrera dans le réacteur à une température extrêmement faible, puisqu'il sera stocké sous forme liquide, et qui devra en sortir à environ 2 500 °K.

Les problèmes clés qui se posent pour la réalisation d'un réacteur de ce genre sont donc liés principalement d'une part aux températures extrêmement élevées qui doivent régner dans le cœur, d'autre part à la concentration d'énergie qui doit être atteinte.

Le choix du matériau qui doit former le support de l'uranium pose de ce fait de nombreuses difficultés, et un compromis doit être trouvé entre les différentes qualités souhaitables :
— bon comportement en température,
— bon modérateur,
— peu absorbeur de neutrons.

Ce qui conduit, en fait, à une sélection très limitée. Les projets actuellement en cours de développement aux Etats-Unis utilisent le graphite, qui présente l'avantage d'avoir un point de sublimation très élevé, et comme inconvénient, entre autres, de n'être pas absolument passif en présence d'hydrogène à haute température. L'emploi d'autres matériaux tels que certains carbures devrait permettre d'atteindre des températures de cœur supérieures à celles qui sont permises par le graphite; toutefois, les projets que l'on peut espérer réaliser en ce sens ne sont certainement pas sans limite et il est peu probable que l'on puisse, avec ce type de réacteur, obtenir des vitesses d'éjection supérieures à 10 000 ou 12 000 m/s (cette dernière correspondant à une température de cœur de l'ordre de 4 000 °K au moins).

Pour aller au-delà, il faut envisager un porteur d'uranium en phase liquide ou même gazeuse. Des études dans ce sens sont actuellement menées aux Etats-Unis et peut-être ailleurs, mais des difficultés considérables de tous ordres sont encore à surmonter avant d'aboutir à une réalisation pratique.

les « Kiwi » et « Nerva »

Après de longues études, les Etats-Unis concentrent actuellement leurs efforts sur deux projets patronnés par l'Atomic Energy Commission.
Il est certainement encore prématuré d'envisager d'effectuer des missions à l'aide de fusées à énergie nucléaire et, *a fortiori*, de prévoir quand des lanceurs de ce genre pourraient être prêts à « prendre la relève » des gros lanceurs à énergie chimique. La série des réacteurs « Kiwi » qui a fait l'objet d'études ces dernières années aux Etats-Unis a permis de faire progresser considérablement la technologie de ces réacteurs; en complément, un projet de réacteur embarqué « Nerva » est actuellement en cours d'étude. On peut espérer qu'avant la fin de cette décennie la phase des essais au sol de « Nerva », comme étage supérieur d'un « Saturn », pourra être atteinte.
Ces deux projets seront revus plus en détail dans le chapitre consacré aux industries nucléaires.
A notre connaissance, aucun projet du même ordre n'est actuellement en étude ou en cours de développement en Europe. Il serait surprenant, toutefois, que dans l'ensemble des programmes nationaux ou européens en cours de discussions un projet de fusée nucléaire ne vît pas le jour.

Phot. U. S. I. S.

les fusées électrothermiques

L'emploi d'un réacteur nucléaire permet d'emmagasiner au sein d'un fluide porteur une quantité d'énergie plus grande que celle qui peut l'être si l'énergie est libérée par une réaction chimique. Toutefois, nous avons vu que cette quantité d'énergie était elle-même limitée par des considérations de tenue mécanique du réacteur, et qu'en fait avec les réacteurs à cœur solide, qui sont les seuls que l'on sache faire aujourd'hui, il n'était guère possible pour le moment de dépasser une température de l'ordre de 3 000 °K. Si l'on souhaite augmenter la quantité d'énergie stockée au sein du fluide, on est donc obligé de procéder autrement. Il est alors nécessaire de passer par l'intermédiaire de l'énergie électrique : si l'on fait jaillir un arc entre deux électrodes et si l'on fait passer le fluide moteur au travers de cet arc, il joue le rôle d'un conducteur dans lequel l'énergie électrique fournie par le générateur est dépensée par effet Joule.

Il est possible d'obtenir ainsi des températures extrêmement élevées : 10 000, voire 20 000 °K et, par conséquent, des impulsions spécifiques sensiblement supérieures à celles qui peuvent être obtenues en utilisant un réacteur nucléaire; toutefois, un certain nombre de considérations interviennent qui, en fait, limitent considérablement l'intérêt de ce genre de solution et ses applications pratiques.

le générateur d'énergie électrique

Pour disposer d'énergie électrique à bord, il est nécessaire de disposer également d'une source première d'énergie, chimique ou nucléaire, d'une machine thermique et d'une génératrice de courant, c'est-à-dire d'un ensemble relativement lourd, et d'autant plus lourd que la puissance à fournir sera plus élevée. A titre d'exemple, si l'on désirait obtenir une fusée de 100 tonnes de poussée et une vitesse d'éjection de 10 000 m/s, on a vu plus haut que la **puissance à fournir était de l'ordre de 5 000 000 de kilowatts.** On conçoit volontiers qu'il n'est pas possible d'embarquer à bord d'une fusée la machine thermique et l'alternateur correspondant à une telle puissance. Pratiquement, on sait faire, ou du moins on espère savoir faire au cours des années à venir, des ensembles générateurs d'énergie électrique pesant, selon la puissance fournie, de 10 à 100 kg par kilowatt électrique.

Pour une puissance d'une dizaine de kilowatts par exemple, il est donc raisonnable de tabler sur un poids de générateur de l'ordre de 500 kg. Compte tenu des différents rendements qui affectent la transformation de l'énergie électrique en énergie cinétique dans le jet, il est également raisonnable d'admettre que la puissance instantanée dans le jet est de l'ordre de 5 kW pour une vitesse d'éjection de 10 000 m/s. Il en résulte donc une poussée de 0,1 kg nettement inférieure au poids du générateur seul, et, *a fortiori,* au poids total du véhicule; pour une vitesse d'éjection plus grande, on obtiendra une poussée encore plus faible.

Il n'est donc pas possible avec un propulseur de ce genre de faire décoller un véhicule depuis le sol, et son emploi ne peut être envisagé que pour des missions pour lesquelles une valeur faible de la poussée n'est pas nécessairement rédhibitoire, pour des étages supérieurs ou des transferts d'orbite par exemple.

où on est limité par les lois de la thermodynamique

Bien qu'on puisse emmagasiner de cette façon dans le fluide moteur une quantité d'énergie très grande, celle qui peut pratiquement être récupérée et transformée en énergie cinétique est limitée. En effet, lorsqu'on chauffe un gaz, de l'hydrogène par exemple, l'énergie est d'abord emmagasinée sous forme d'énergie de température correspondant aux différents mouvements d'agitation des atomes constituant la molécule. A partir d'une certaine température variable avec la pression, mais disons de l'ordre de 2 500 °K, une partie importante de l'énergie est utilisée à dissocier les molécules en leurs atomes. A une température supérieure, de l'ordre de 5 000 à 6 000 °K, ce sont les atomes qui sont eux-mêmes dissociés (on dit aussi ionisés) :

$$H = H^+ + e^-.$$

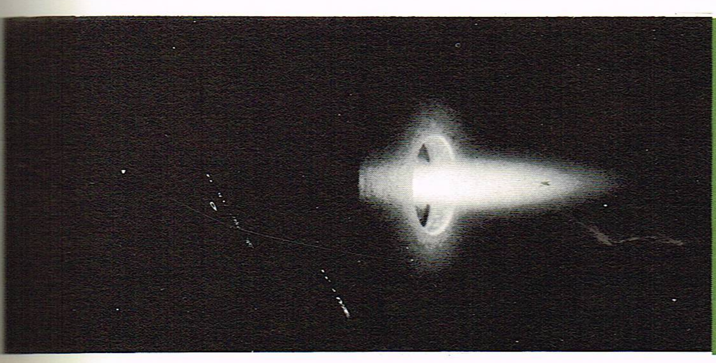
L'expérience montre qu'au cours de la détente les énergies d'ionisation et de dissociation ne sont pratiquement pas récupérées, la composition du mélange des gaz demeurant pratiquement la même au cours de celle-ci.

Il s'ensuit que si l'on désire disposer d'une énergie extrêmement grande dans les gaz éjectés, il sera nécessaire de fournir sous forme électrique une énergie encore beaucoup plus grande, dont une grande proportion le sera en pure perte. Pour cette raison, dans les propulseurs à hydrogène par exemple, on n'envisage pratiquement jamais de chauffer celui-ci à des températures supérieures à 3 500 ou 4 000 °K.

les matériaux

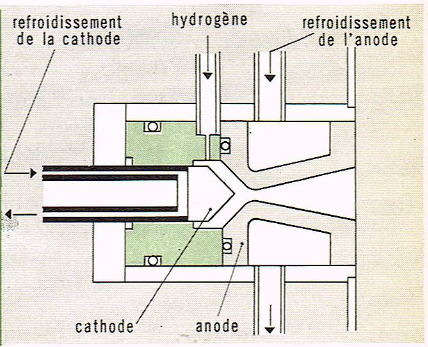
Enfin, bien que présentant moins d'acuité que dans le cas du réacteur nucléaire, le problème de la bonne tenue des matériaux n'en existe pas moins dans ce genre de fusée : tenue des électrodes (généralement en tungstène), comportement de la tuyère nécessairement non refroidie.

Aucune expérimentation en vol n'a encore été faite avec des propulseurs de ce genre, probablement parce qu'on ne dispose pas de générateur d'énergie prêt à fonctionner dans des conditions acceptables. De nombreuses fusées expérimentales fonctionnent toutefois en laboratoire et, le cas échéant, seraient certainement prêtes à être utilisées.



Photographie d'un moteur électrothermique à arc et schéma de fonctionnement.

Dans la plupart des configurations envisagées, l'arc jaillit entre une électrode centrale en tungstène et la tuyère annulaire, l'hydrogène étant injecté autour de l'électrode centrale. Dans certains cas, un système permettant de faire avancer l'électrode, pour compenser son usure, peut être prévu. Généralement, on utilise du courant continu; des systèmes peuvent être envisagés permettant d'utiliser du courant bi- ou triphasé.



les fusées électriques

Pour accélérer une masse de gaz par détente, il faut lui communiquer au préalable une certaine énergie, qui sera transformée au cours de la détente en énergie cinétique. On a vu que la quantité d'énergie qui peut être ainsi communiquée à la masse de gaz en question est dans tous les cas plus ou moins limitée, et que, par conséquent, il en est de même des vitesses que l'on peut espérer atteindre en opérant de cette façon.

Le principe de la propulsion électrique est de rendre les molécules des gaz ou, plus généralement, les particules que l'on se propose d'éjecter, sensibles à l'action d'un champ magnétique ou électrique et de leur communiquer de l'énergie cinétique par l'intermédiaire de ce champ.

Il est possible d'obtenir de cette façon des vitesses beaucoup plus élevées que celles qui sont réalisées en passant par l'intermédiaire d'un chauffage; selon le dispositif adopté, il est relativement aisé d'obtenir des vitesses d'éjection de l'ordre de 50 000, 100 000 m/s, voire supérieures.

La seule méthode que l'on connaisse pour rendre un gaz suffisamment conducteur pour qu'il soit sensible à l'action d'un champ est d'arracher à chaque atome un ou plusieurs électrons (ionisation) : on obtient ainsi à l'état libre des charges négatives constituées par les électrons eux-mêmes et des charges positives. Dans une fusée électrique, une partie de l'énergie à fournir doit donc l'être initialement pour ioniser le fluide moteur : elle ne peut être par la suite transformée en énergie cinétique.

Q étant la charge de l'ion de masse *m* et E le potentiel accélérateur. L'ionisation se faisant toujours par perte d'un électron, Q est donc égale à e (en valeur absolue).

Le cæsium est souvent choisi, notamment pour la facilité avec laquelle il peut être ionisé.

Les principaux problèmes posés par la réalisation d'une telle fusée sont :

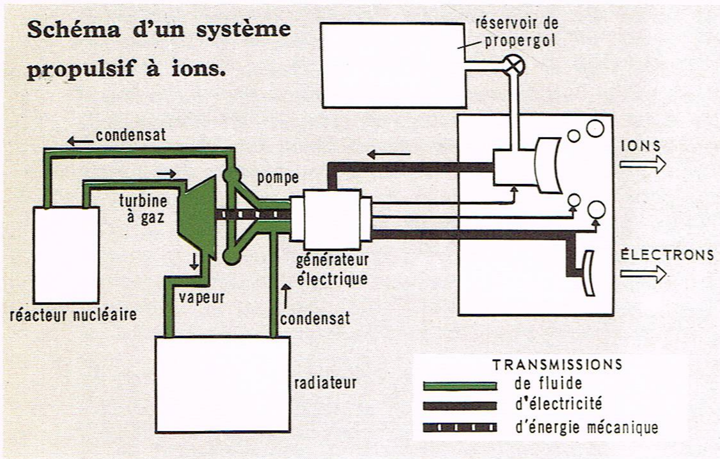
— la production des ions avec un bon rendement au travers d'une électrode poreuse, en général en tungstène;

— la résistance de l'électrode accélératrice, qui est, en fait, soumise à l'impact des ions qui, pour des raisons diverses, n'ont pas suivi une trajectoire convenable;

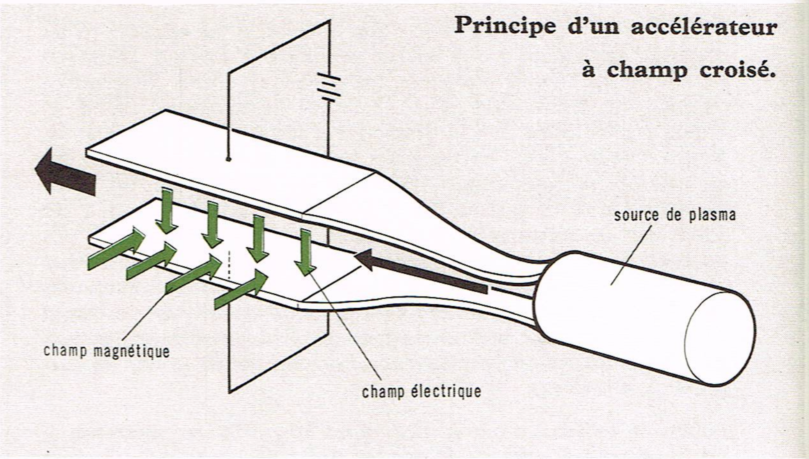
— la neutralisation du jet à la sortie du propulseur : si le sillage demeure chargé positivement, l'éjection des ions positifs deviendra impossible. D'autre part, si les électrons provenant de la séparation des charges ne sont pas évacués hors du véhicule, celui-ci se chargera négativement et le fonctionnement du moteur deviendra aussi impossible; il est donc nécessaire de réinjecter le flux d'électrons dans le jet d'ions à la sortie du propulseur en neutralisant celui-ci.

Les vitesses d'éjection qui peuvent être atteintes ne sont guère limitées en valeurs supérieures; il est possible, *a priori*, d'envisager des vitesses de plusieurs centaines de milliers de mètres/seconde; par contre, en raison du phénomène des charges

Schéma d'un système propulsif à ions.



Principe d'un accélérateur à champ croisé.



Les fusées électriques peuvent être groupées en deux familles :
— celles dans lesquelles l'accélération se produit dans un champ électrostatique (fusées ioniques);
— celles dans lesquelles l'accélération se produit dans un champ électromagnétique (fusées à plasma).

fusées ioniques...

Si l'on sait ioniser parfaitement un milieu et séparer les charges positives et les charges négatives, il est possible d'accélérer les unes ou les autres dans un champ électrostatique. Dans la fusée ionique, on procède au préalable à cette séparation, et on accélère les ions ainsi obtenus entre deux électrodes, comme il est indiqué sur le schéma ci-dessus.

La vitesse prise par ces ions est exprimée par la formule

$$V = \sqrt{2 \frac{Q}{m} \times E},$$

d'espace entre les deux électrodes, elles sont limitées en valeurs inférieures, et pratiquement il est difficile de descendre audessous de 50 000 m/s avec un fonctionnement convenable.

...et propulseurs à éjection de plasma

On appelle « plasma » un gaz ionisé dont la densité moyenne de charge est nulle (quelle que soit la proportion d'atomes ionisés). De nombreux dispositifs peuvent être envisagés pour produire et accélérer un plasma. Ils sont tous fondés sur le même principe qui consiste à accélérer dans un champ électromagnétique un gaz préalablement ionisé. Dans tous les cas, l'ionisation est effectuée par chauffage, soit par arc, soit par induction, l'accélération étant due à un champ électromagnétique créé extérieurement ou engendré par le plasma lui-même.

Nous nous limiterons à décrire quelques types d'accélérateurs de plasma, qui, semble-t-il, sont assez bien adaptés à l'utilisation comme propulseurs et ont fait l'objet de réalisations en ce sens.

fonctionnement continu...

Accélérateur à champs croisés :

Un gaz est ionisé au travers d'un arc, puis accéléré dans le champ de force résultant d'un champ électromagnétique et d'un champ électrostatique croisé (v. *fig.* ci-contre). Le mouvement pris par les particules chargées est perpendiculaire au plan formé par les directions de ces deux champs.

Tube à onde progressive :

Un champ non uniforme et mobile est produit dans un tube contenant le plasma. Ce champ est engendré par des enroulements alimentés par des courants polyphasés ou par des lignes à retard.

...ou discontinu :

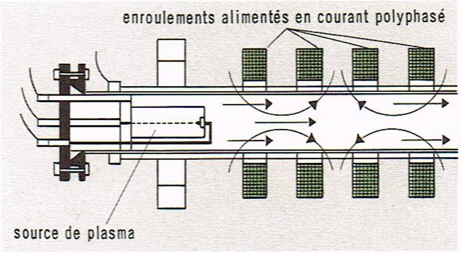
A la différence des systèmes à fonctionnement continu, la plupart des appareils à fonctionnement intermittent utilisent le champ magnétique propre de la décharge pour accélérer une masse limitée de plasma. Ils ne comportent pas, en général, d'enroulements importants destinés à créer le champ, mais ils nécessitent la présence d'une batterie de condensateurs pour stocker l'énergie entre deux décharges. Dans le cas de propulseurs de faible puissance, l'emploi d'un tel générateur permet de travailler sur des quantités de plasma beaucoup plus grandes si on le souhaite que celles qui sont imposées par un fonctionnement continu ; de ce fait, on n'est pas conduit à des dimensions de propulseur extrêmement petites, cause éventuelle de mauvais fonctionnement et de mauvais rendement.

Des résultats expérimentaux ont été obtenus avec plusieurs systèmes de fonctionnement continu ou intermittent, en particulier dans les laboratoires américains. Comme les propulseurs à ions, les propulseurs à accélération de plasma sont encore, pour la plupart, des appareillages de laboratoire dont l'expérimentation en vol ne pourra pas être envisagée tant que l'on ne disposera pas des générateurs d'énergie nécessaires.

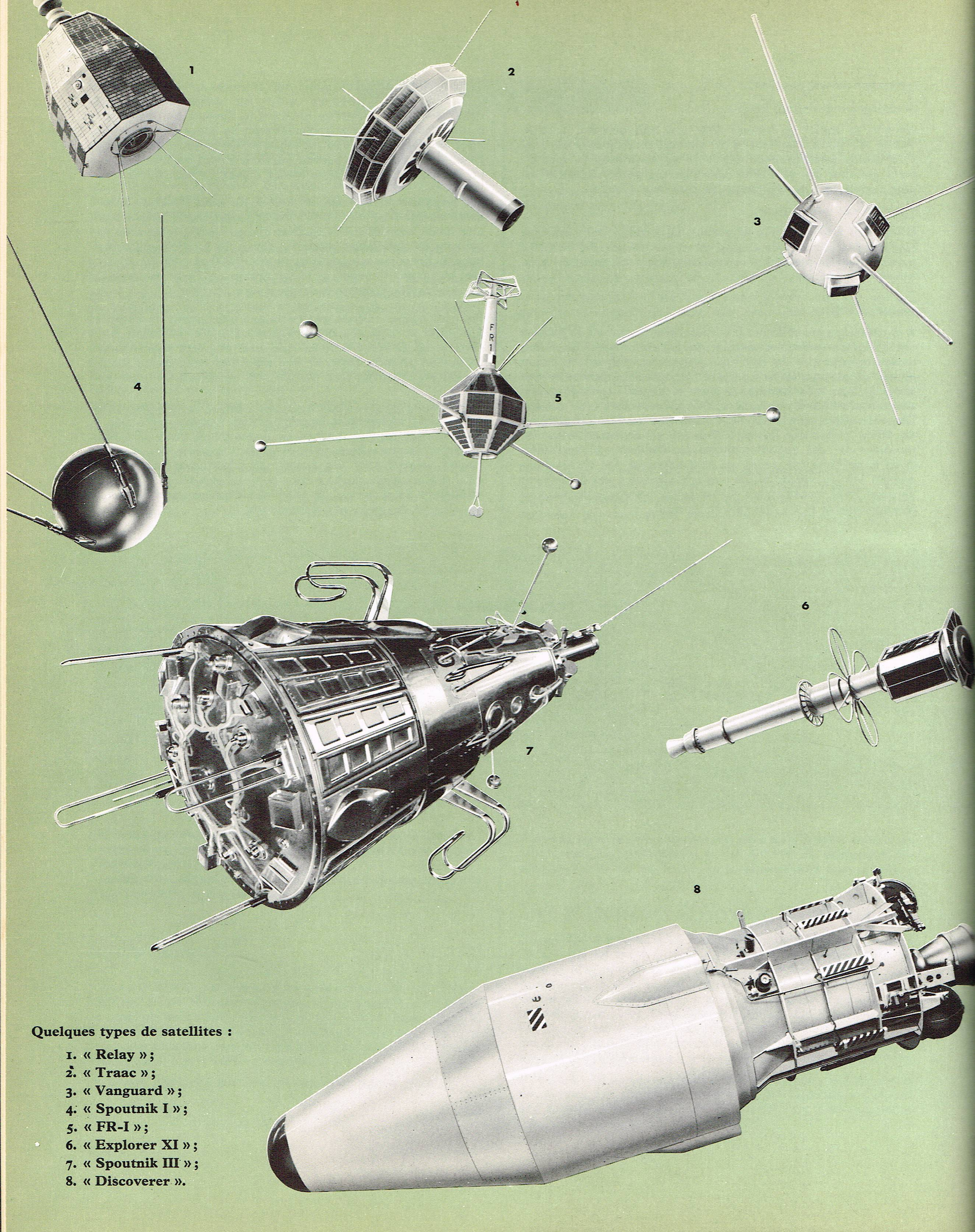
Les processus fondamentaux mis en jeu sont, de plus, encore assez mal connus, et il est difficile, entre autres, de fixer *a priori* une limite aux vitesses d'éjection qui peuvent être atteintes en opérant de cette façon. Avec des dispositifs fonctionnant actuellement, des vitesses comprises entre 10 000 et 100 000 m/s ont été obtenues. Certains auteurs prétendent que des valeurs beaucoup plus grandes pourraient être atteintes.

Pour les mêmes raisons que celles que nous avons développées plus haut à l'occasion des propulseurs à arc, les fusées électriques ne peuvent fournir que des poussées extrêmement faibles ne pouvant convenir que pour des missions de longue durée, pratiquement sans travail dans un champ de pesanteur.

Indépendamment de l'absence de générateur d'énergie convenable, les missions actuellement envisagées dans l'espace Terre-Lune ne semblent pas justifier l'emploi de fusées de ce genre. Toutefois, lorsqu'on voudra envoyer vers Vénus ou vers Mars des vaisseaux habités, compte tenu des vitesses caractéristiques considérables qu'il faudra communiquer à ceux-ci, il est possible qu'il se révèle intéressant de recourir à la propulsion électrique afin de réduire les charges de propergols à transporter.

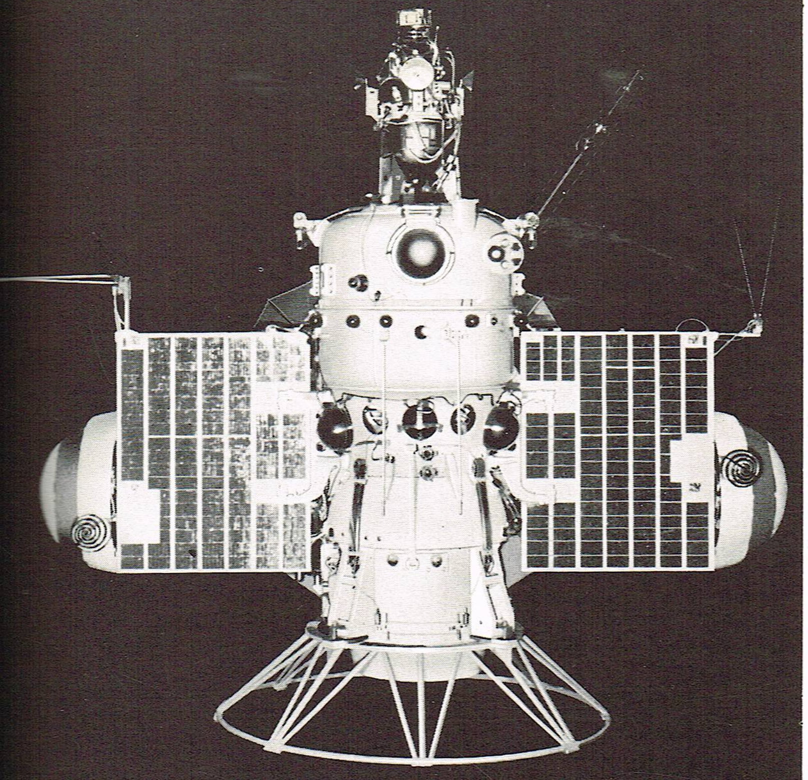


Tube à onde progressive.



Quelques types de satellites :

1. « Relay » ;
2. « Traac » ;
3. « Vanguard » ;
4. « Spoutnik I » ;
5. « FR-I » ;
6. « Explorer XI » ;
7. « Spoutnik III » ;
8. « Discoverer ».



La sonde spatiale « Mars I ».



Le satellite passif « Echo II » est lancé replié et gonflé après mise sur orbite.

satellites

et véhicules spatiaux

Si nous savons lancer une masse avec une vitesse initiale suffisamment élevée (première vitesse spatiale) par rapport à la Terre, cette masse ne retombe pas sur la Terre, mais tourne autour d'elle suivant une trajectoire elliptique ou circulaire. Nous avons obtenu un satellite artificiel de la Terre : « Telstar », par exemple.

Si nous savons communiquer à cette masse une vitesse supérieure à la vitesse de libération, elle va, en théorie, s'écarter indéfiniment de la Terre. En réalité, la Terre n'est pas le seul centre d'attraction de l'univers, et dès que la masse sera suffisamment éloignée de la Terre, l'attraction du Soleil, en particulier, deviendra prépondérante. Si la vitesse communiquée à la masse ne dépasse pas la « vitesse de libération par rapport au Soleil », ce qui est le cas en général, la masse va devenir un satellite du Soleil. On appelle « sonde spatiale » une telle planète artificielle : « Mariner », par exemple.

Naturellement, si l'on savait communiquer à une telle sonde une vitesse supérieure à la vitesse de libération par rapport au Soleil, la sonde s'écarterait indéfiniment du Soleil et quitterait le système solaire pour se diriger vers les étoiles plus ou moins proches.

Il est possible également de créer des satellites de corps célestes autres que la Terre ou le Soleil. Supposons, par exemple, que nous lancions un satellite artificiel de la Terre dont la trajectoire soit très excentrique et l'amène au voisinage de la Lune. La Lune n'ayant pas un champ de pesanteur très intense, le satellite de la Terre aura généralement une vitesse supérieure à la vitesse de libération par rapport à la Lune; la Lune ne pourra donc le capter et s'en faire un satellite propre, elle ne fera que déformer la trajectoire, mais le satellite reviendra au voisinage de la Terre. Exemple : « Lunik II », qui a photographié la face jadis inconnue de la Lune.

Par contre, si au voisinage de la Lune on sait freiner suffisamment le satellite terrestre pour que sa vitesse devienne inférieure à la vitesse de libération par rapport à la Lune, cet ancien satellite de la Terre deviendra un satellite de la Lune et ne reviendra plus autour de la Terre (sous réserve que la trajectoire par rapport à la Lune ne soit pas trop excentrique en direction de la Terre).

La même manœuvre peut naturellement être répétée pour une sonde spatiale lorsqu'elle passe au voisinage d'une planète; on créerait ainsi un satellite de Mars ou de Vénus par exemple.

Les fusées, dont le fonctionnement vient d'être étudié, fournissent le moyen de propulsion permettant de réaliser de telles missions. *Cependant, aussi intéressante que puisse être la simple création de satellites artificiels de la Terre, du Soleil, ou des planètes, celle-ci n'est pas une fin de la recherche spatiale. Pour pouvoir procéder à des expériences plus complètes, il faut aussi que le satellite*

lui-même devienne un véritable laboratoire en miniature, ce qui nécessite la solution de nouveaux problèmes techniques parfois fort complexes. C'est à la description sommaire de quelques-unes des missions ainsi réalisées et à l'énoncé des principaux problèmes techniques posés que seront consacrées les pages suivantes.

Nous ne parlerons pas ici des applications militaires possibles des véhicules spatiaux. Les missions se séparent alors en deux catégories simples :

a) les missions « scientifiques »,

b) les missions « utilitaires » (satellites météorologiques, de navigation, de télécommunications),

bien que certaines missions dites « utilitaires » puissent apporter de précieux renseignements à la science.

l'exploration de l'espace

Astronomie et géophysique sont entrées, depuis 1957, date des premiers progrès de la technique spatiale, dans une ère de développement extrêmement rapide.

En astronomie, par exemple, la possibilité d'emporter des instruments au-delà de l'atmosphère terrestre permet d'étendre considérablement le champ d'observation. De nombreux rayonnements sont, en effet, absorbés par l'atmosphère; ce qui interdit, au sol, l'étude spectrographique dans de vastes zones du domaine des fréquences. La mise sur orbite de satellites (ou simplement l'envoi de sondes spatiales) au voisinage des planètes proches permet également de recueillir des renseignements qui ne seraient sans doute pas disponibles autrement. En géophysique, l'emploi systématique des satellites s'est révélé un moyen d'investigation remarquable, que ce soit pour l'étude détaillée de l'atmosphère, même élevée, ou simplement pour connaître avec précision la forme de la Terre.

Il n'est pas possible de passer en revue ici toutes les expériences scientifiques qui ont été effectuées à ce jour au moyen de satellites ou de sondes spatiales. Un inventaire complet des moyens techniques mis en œuvre pour ces expériences serait également long, d'autant plus qu'un appareillage assez complexe et varié a été utilisé; appareillage souvent très spécifique d'ailleurs et qui relève davantage de la technique du laboratoire de physique que du domaine purement spatial. Nous nous contenterons de citer ici un exemple typique de résultats obtenus grâce aux recherches spatiales : celui des ceintures de particules entourant la Terre.

satellites et ceinture de Van Allen

Les mesures effectuées au moyen des tout premiers satellites (« Spoutnik » et « Explorer ») ont révélé qu'à l'exclusion des régions polaires la Terre était entourée d'un véritable nuage de particules électrisées. Il s'agit d'électrons et de protons « captés » par le champ magnétique terrestre, astreints à s'enrouler autour des lignes de force de ce champ et à rebondir alternativement d'un hémisphère à l'autre (phénomène de réflexion, dû au fait que le champ magnétique se concentre au voisinage des régions polaires).

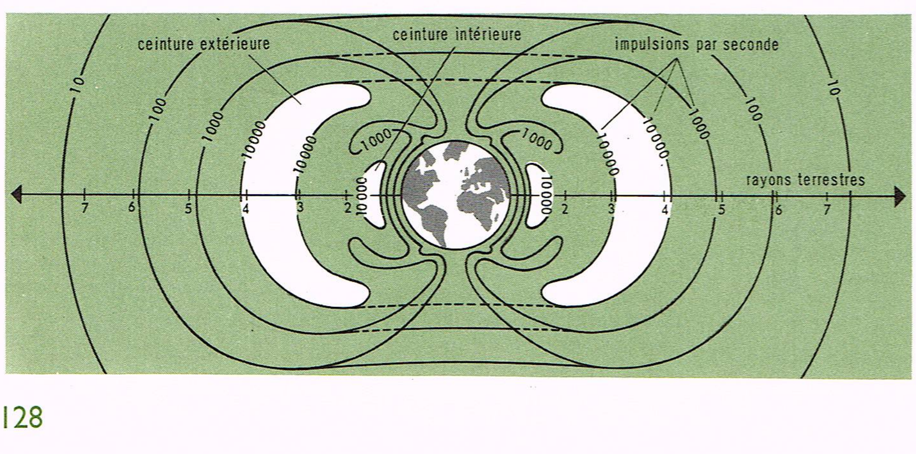
Les particules les plus énergétiques restent au voisinage de la Terre, où le champ est plus intense; quant à celles de faible énergie, elles peuvent s'éloigner jusqu'à 10 rayons terrestres (environ 60 000 km), sans pour cela échapper à l'attraction du champ magnétique. Il en résulte un anneau de particules, entourant la Terre, auquel on a donné le nom de « ceinture de Van Allen ».

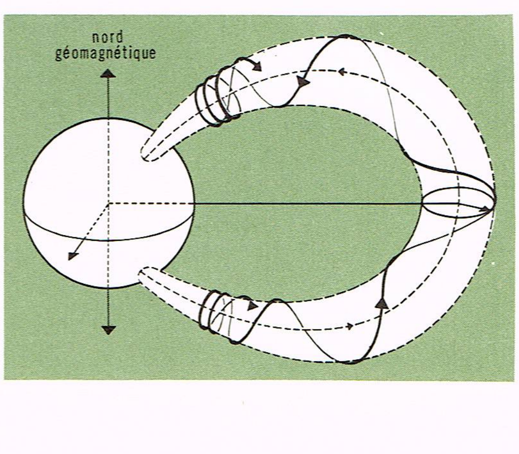
On constate que le flux peut atteindre 10^8 particules par centimètre carré et par seconde, c'est-à-dire que cent millions de protons ou d'électrons traversent toutes les secondes une surface de un centimètre carré. Cela provient de ce que les particules très énergétiques ont des vitesses voisines de celles de la lumière : ce flux correspond, en fait, à une densité très faible (10^{-2} à 10^{-3} particule par centimètre cube) et à un poids total de particules pour l'ensemble de la ceinture de... quelques grammes seulement (et non de millions de tonnes!).

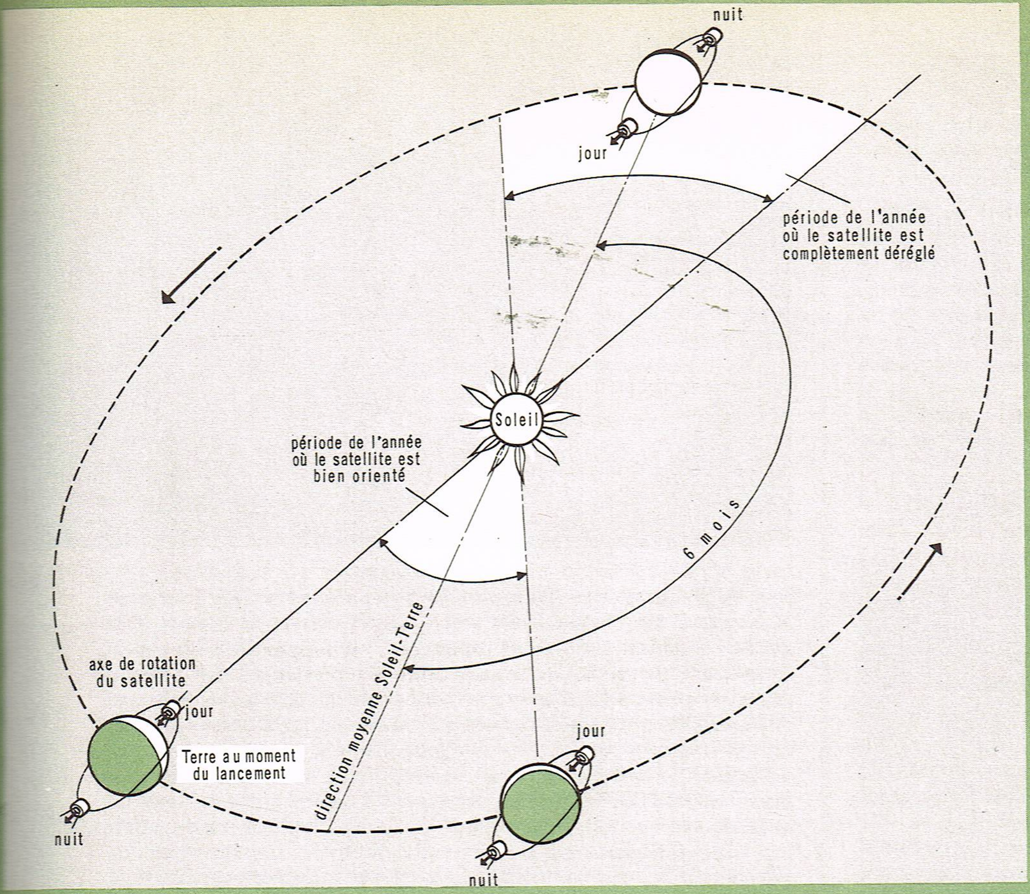
La ceinture de Van Allen peut ainsi ne paraître qu'un accident minime à l'échelle spatiale, cependant sa présence obligera à prévoir une protection des futurs astronautes contre les rayonnements auxquels ils seront soumis pendant sa traversée : après quelques mois de fonctionnement parfait, le premier satellite « Telstar » s'est arrêté d'émettre à la suite des dégâts provoqués dans son appareillage électronique par un séjour prolongé dans la ceinture de Van Allen.

les satellites météorologiques

On classe sous ce nom les satellites destinés à prendre des photos des formations nuageuses entourant la Terre et aussi à mesurer le rayonnement infrarouge émis par la surface terrestre. Ces informations devraient, en effet, permettre d'améliorer sensiblement les prévisions météorologiques et éventuellement de

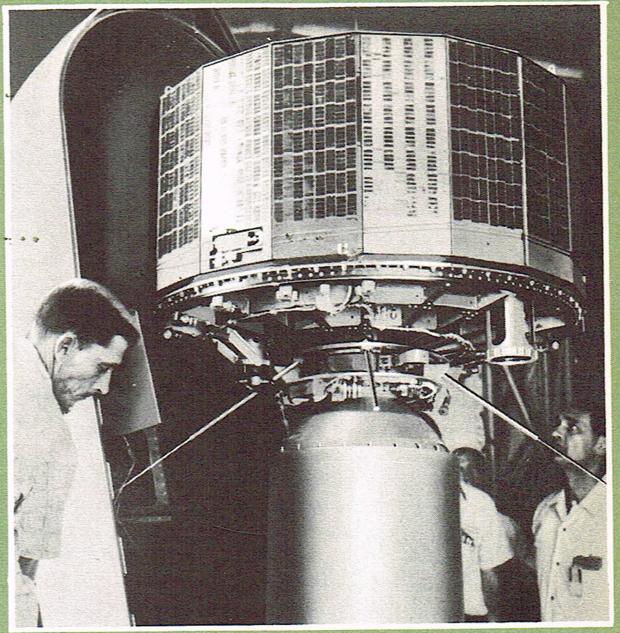






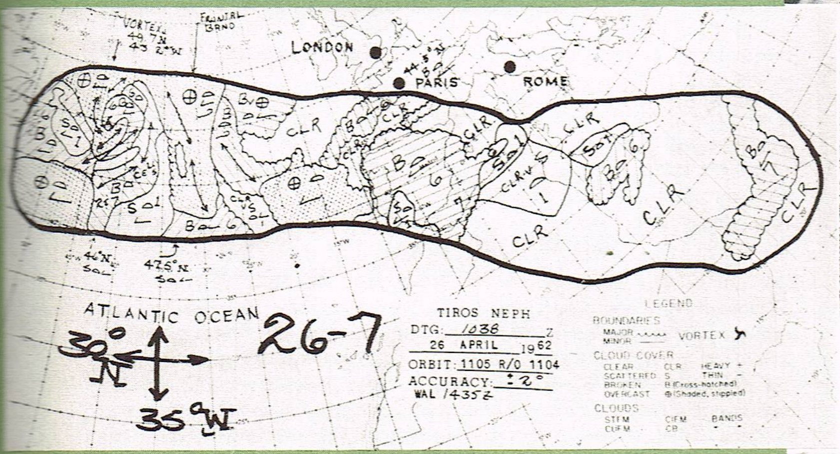
Pointage à différentes époques de l'année.

Le satellite météorologique « Tiros ».



Montage sur la fusée lanceuse.

Ci-dessous, carte météorologique établie d'après les informations fournies par les photographies que transmettent ces satellites.





Exemple de photographie transmise par le satellite « Tiros IV », montrant une formation cyclonique.

Phot. U. S. I. S.

conduire à une meilleure compréhension des phénomènes régissant l'évolution du temps. Ces satellites sont de deux types : « Tiros » et « Nimbus ».

« Tiros » I a été lancé le 1[er] avril 1960 sur une orbite de 690 km de périgée et 750 km d'apogée, de 48° d'inclinaison et de période voisine de 99 minutes. C'est un satellite cylindrique de 48 cm de haut et 107 cm de diamètre, qui pèse environ 123 kg. Ses organes essentiels sont deux caméras de télévision : la première, de grande ouverture, photographie une zone de 1 125 km de diamètre; la seconde, de faible ouverture, prend des photos plus détaillées d'une zone de 115 km de diamètre. Ces images de télévision sont enregistrées à l'aide d'un magnétophone et retransmises à une station au sol lorsque « Tiros » survole les Etats-Unis.

Pour qu'un satellite puisse prendre des photos de la Terre, il faut d'une part que les caméras soient braquées en direction de la Terre et de l'autre que la Terre soit éclairée par le Soleil. On communique au satellite une rotation de vitesse angulaire 12 tours-minute environ et d'axe défini par la direction moyenne d'une droite Soleil-Terre pendant la durée de vie utile du satellite. Cela permet aux caméras de rester précisément dirigées vers la Terre, du moins aux moments propices (quand le satellite, par contre, survole une zone où il fait nuit, ses caméras ne sont pas braquées vers la Terre, mais vers le Ciel).

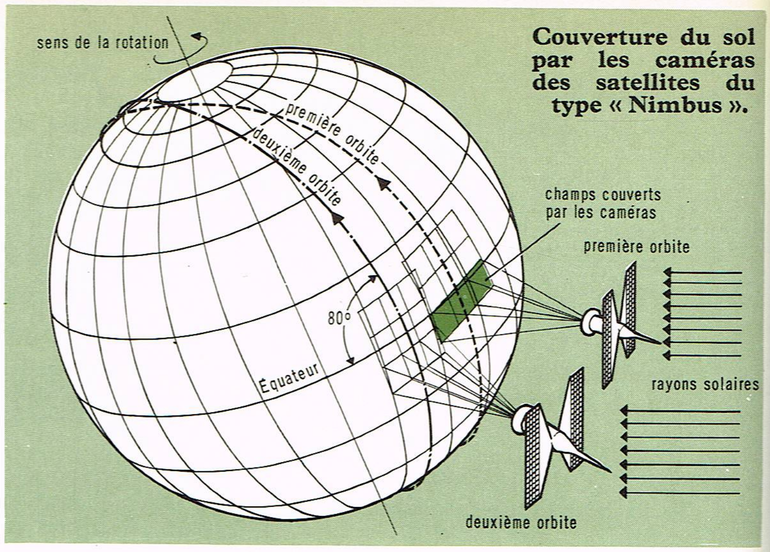
Puisque la direction Soleil-Terre change au cours du temps, ce réglage ne peut être efficace que pendant deux à trois mois au maximum. Un tel mode de pointage est par contre très simple et était bien adapté à « Tiros », qui était un satellite expérimental. (On voit notamment [page précéd.] que six mois après le lancement les caméras sont pointées vers la Terre systématiquement lorsqu'il fait nuit et sont pointées vers le Soleil lorsque le satellite survole une région éclairée.)

Le satellite « Nimbus », par contre, est beaucoup plus évolué (mais aussi sensiblement plus lourd : 300 kg au lieu de 123). Il doit être lancé sur une orbite polaire et nécessite l'action de petites fusées auxiliaires pour rester correctement pointé; de plus, le champ des caméras est tel qu'entre deux passages du satellite la Terre ait tourné d'un angle égal au champ des caméras. On obtient ainsi une couverture complète de la face éclairée de la Terre. Des satellites de ce type permettent d'étudier, sur l'ensemble de la Terre, l'évolution des formations nuageuses; ils ont notamment mis en évidence l'organisation à grande échelle des nuages, manifestée par exemple par des formations en spirale qui peuvent s'étendre sur des distances atteignant 2 500 km. L'existence de tels systèmes nuageux montre bien l'intérêt qu'il peut y avoir à utiliser des observations instantanées très étendues pour la prévision du temps.

satellites de navigation et système Transit

Lorsqu'un bateau désire connaître avec précision sa position au cours d'une traversée, il est nécessaire de procéder à des observations astronomiques. En supposant que l'on connaisse l'heure avec exactitude, on peut en effet trouver sur des tables calculées à l'avance la position de certains astres déterminés; il est alors possible de déduire des directions sous lesquelles ces astres sont observés quelle est la position du point d'observation.

Une telle mesure n'est naturellement réalisable que si le ciel est suffisamment dégagé pour que l'on puisse voir les astres choisis. Il est possible de se débarrasser de cette condition si l'objet observé émet des signaux perceptibles même à travers les nuages, d'où l'idée d'utiliser non des objets célestes, mais des satellites contenant un émetteur de radio. Le seul problème qui reste à résoudre est maintenant le suivant : connaissant en permanence le temps d'une part et l'orbite du satellite d'autre part, comment peut-on déduire sa position sur la Terre de la simple analyse de signaux de radio émis par le satellite? La solution est simple, au moins théoriquement, et repose sur l'utilisation de l'effet Doppler-Fizeau : lorsqu'un observateur, supposé fixe, capte les ondes provenant d'un émetteur mobile, la fréquence à l'émission étant constante, la fréquence du signal reçu varie au cours du temps et ne dépend que de la vitesse relative entre l'émetteur et l'observateur. En particulier, si un émetteur de fréquence fixe est installé à bord d'un satellite, un observateur situé au sol détecte une onde de fréquence variable. La courbe donnant la fréquence à la réception fixe ainsi, en fonction du temps, la vitesse relative entre le satellite et l'observateur. Si l'observateur connaît la trajectoire du satellite, il peut, à partir de la connaissance de ces vitesses relatives, déterminer sa position par rapport à la trajectoire du satellite.



Couverture du sol par les caméras des satellites du type « Nimbus ».

La trajectoire d'un satellite artificiel se modifie progressivement au cours du temps (à cause des diverses perturbations, comme par exemple la traînée due à l'atmosphère même très raréfiée, le champ magnétique terrestre, l'aplatissement de la Terre, etc.). Dans le système de navigation utilisant les satellites « Transit », des stations de repérage au sol déterminent avec une grande précision la trajectoire actuelle du satellite (forme et position de l'orbite par rapport à la Terre). Lors du passage du satellite au voisinage d'une telle station, ces informations sont émises en direction du satellite, où, après réception, elles sont enregistrées à l'aide d'un magnétophone.

Le satellite contient en outre deux émetteurs : l'un, de grande précision, fournit une onde de fréquence constante au cours du temps; l'autre émet simplement les renseignements sur la trajectoire, enregistrés sur le magnétophone.

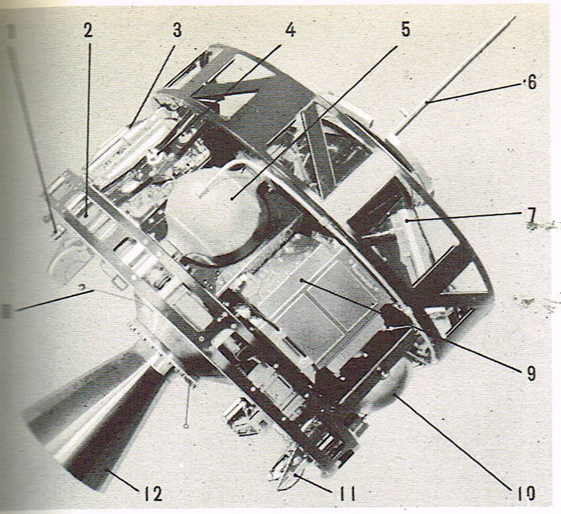
Un observateur au sol, grâce à cette seconde émission, connaît à chaque passage les caractérisques exactes de la trajectoire du satellite. L'onde qu'il reçoit de l'émetteur « étalon » lui apparaît comme étant de fréquence variable à cause de l'effet Doppler. Comme il a été dit plus haut, il peut alors en déduire sa position par rapport à l'orbite du satellite; puisqu'il connaît, de plus, avec précision la position de cette orbite par rapport à la Terre, il en déduit sa propre position géographique.

Grâce à l'artifice consistant à faire émettre par le satellite lui-même toutes les caractéristiques de sa propre trajectoire, il est alors possible d'atteindre une assez grande précision, tous les éléments du calcul étant eux-mêmes connus avec grande précision. On espère ainsi avec un tel système permettre à un bateau de connaître, par tous les temps, sa position avec une erreur comprise entre 200 et 400 m!

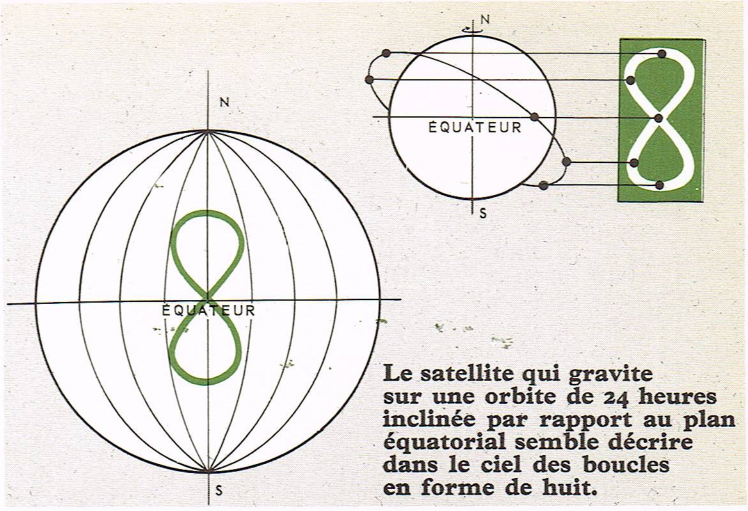
les télécommunications par satellites

Les satellites de télécommunications constituent l'application spatiale la plus directement utilisable pour l'ensemble des usagers. Les lancements de satellites expérimentaux se sont multipliés depuis celui de « Telstar », en 1962, et la mise en place d'un réseau mondial de télécommunications par satellites

Le satellite « Syncom ».



1. Fusée-vernier (orientation);
2. Batteries au nickel-cadmium;
3. Récepteur du relais hertzien;
4. Fusée-vernier (vitesse);
5. Eau oxygénée;
6. Antenne coaxiale;
7. Émetteur du relais hertzien;
8. Antenne de télémétrie et télécommande;
9. Récepteur pour télécommande;
10. Azote;
11. Cellule solaire (orientation);
12. Moteur pour la mise en orbite circulaire.



Le satellite qui gravite sur une orbite de 24 heures inclinée par rapport au plan équatorial semble décrire dans le ciel des boucles en forme de huit.

est actuellement à l'étude. L'intérêt de ce type de satellite provient essentiellement de la saturation actuelle des moyens de télécommunication classiques à grande distance. La seule solution dont on dispose pour augmenter le trafic consiste à utiliser des liaisons supplémentaires par radio, mais celles-ci ne peuvent être alors effectuées que sur des longueurs d'onde assez courtes, toutes les grandes longueurs d'onde étant déjà utilisées. De telles ondes ne se propagent pratiquement qu'en ligne droite et exigent donc que l'émetteur et le récepteur soient en vision directe l'un de l'autre. On sait que ce problème est résolu en pratique par l'emploi de relais successifs, tous en vision directe l'un de l'autre et composés chacun d'un poste émetteur et d'un poste récepteur. Naturellement, la distance entre deux relais successifs peut être d'autant plus grande (et donc le nombre de relais d'autant plus faible) que ceux-ci sont situés à plus grande altitude, d'où l'idée simple de disposer le relais à bord d'un satellite. Ces relais peuvent être réalisés pratiquement selon plusieurs systèmes distincts que nous passons rapidement en revue ci-après.

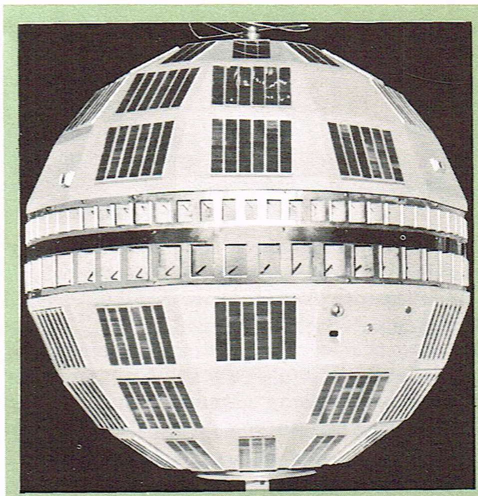
satellites stationnaires

La période de rotation d'un satellite autour de la Terre croît avec l'altitude du satellite; il existe donc une altitude pour laquelle la période est de vingt-quatre heures, c'est-à-dire pour laquelle le satellite tourne dans l'espace à la même vitesse de rotation que la Terre. Si l'orbite du satellite est équatoriale, le satellite reste donc ainsi en permanence au-dessus d'un point fixe de l'équateur (satellite stationnaire); par contre, si l'orbite est inclinée par rapport au plan de l'équateur, la projection du satellite sur la Terre décrit une courbe en huit comme celle qui est représentée ci-contre. L'ensemble de cette courbe est décrit avec une période de vingt-quatre heures.

L'altitude pour laquelle un satellite est stationnaire est malheureusement assez élevée (voisine de 36 000 km) et la réalisation d'une telle mission est une opération complexe, par l'ensemble des manœuvres à effectuer, et qui nécessite par ailleurs l'emploi d'une fusée puissante, bien que la masse utile à transporter sur orbite soit relativement petite (quelques dizaines de kilogrammes pour les satellites expérimentaux actuellement lancés, peut-être quelques centaines de kilogrammes dans l'avenir). En outre, le problème du maintien du satellite au-dessus d'un point de la Terre malgré les diverses perturbations possibles peut être relativement délicat à résoudre correctement. Par contre, grâce à leur altitude élevée, il est possible théoriquement de n'utiliser que trois satellites stationnaires pour « couvrir » l'ensemble de la Terre.

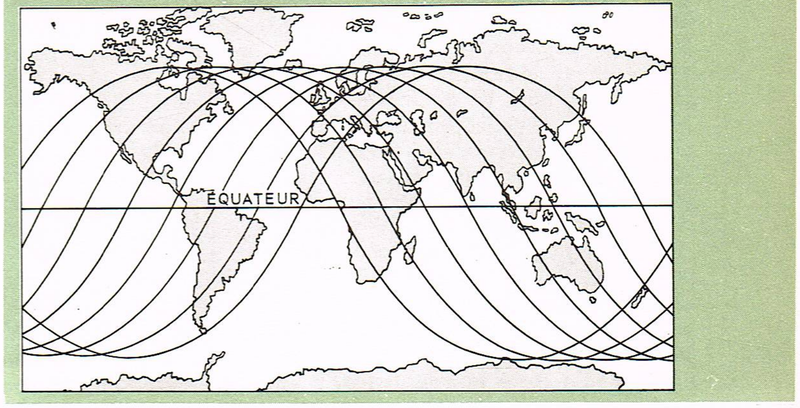
« Telstar » et « Relay »

En dehors du cas particulier des satellites synchrones ou stationnaires, les satellites ne repassent pas au-dessus des mêmes points de la Terre au cours de deux révolutions consécutives; on obtient des satellites dits « à défilement » : c'est le cas de



Le satellite « Telstar ».

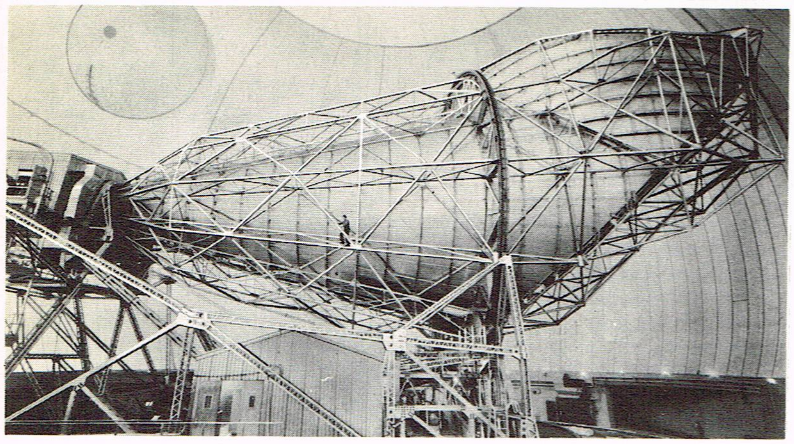
Les satellites de défilement balaient la surface du globe (*ci-dessous*), puisque celui-ci tourne vers l'est entre deux passages consécutifs de l'engin par un même point de son orbite.



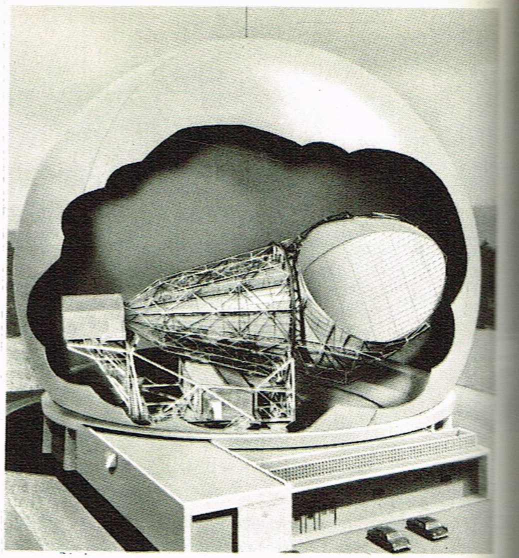
« Telstar » et « Relay ». On dispose, dans ce cas général, d'une très grande liberté dans le choix de l'orbite; l'altitude est seulement choisie suffisamment élevée pour que le satellite ne soit pas freiné par l'atmosphère terrestre et possède ainsi une durée de vie assez grande.

Les satellites à défilement présentent l'avantage de ne pas poser de graves problèmes de lancement et notamment de ne pas nécessiter l'emploi de très grosses fusées. Ils ont, par contre, deux inconvénients importants : en premier lieu, si l'on veut pouvoir transmettre en permanence entre deux points de la Terre, il est nécessaire qu'il y ait toujours un satellite d'où l'on puisse voir à la fois les deux points donnés. Cela exige l'emploi simultané de nombreux satellites (sans doute de l'ordre d'une cinquantaine si l'écart relatif des satellites n'est pas contrôlé); en second lieu, ils nécessitent l'emploi de stations au sol très complexes.

Phot. U. S. I. S.



Ce type d'antenne tournante, d'un poids de 350 tonnes, qui suit constamment le satellite dans sa course pour mieux capter ses émissions, existe en plusieurs exemplaires dans le monde; l'un d'eux est en France, à la station de Pleumeur-Bodou (Côtes-du-Nord).



la station de Pleumeur-Bodou

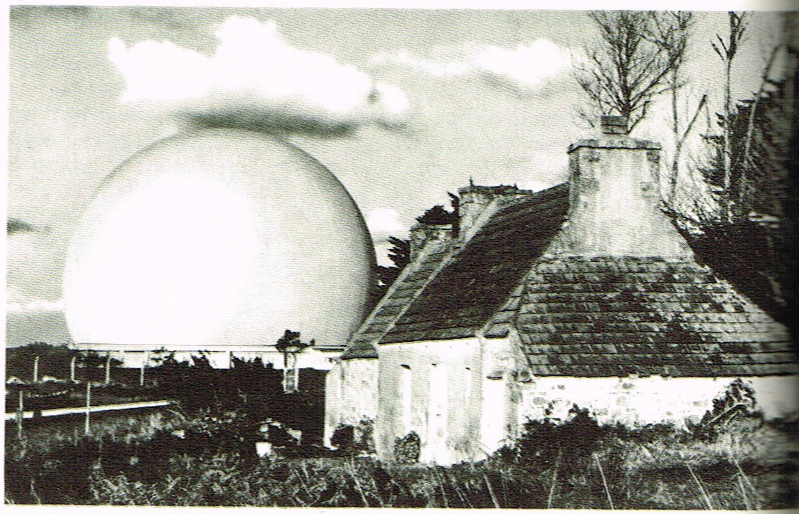
Pendant toute la durée de la transmission, un satellite à défilement se déplace notablement par rapport aux deux stations au sol. Or, comme la puissance électrique disponible à bord du satellite est faible, le poste récepteur disposé à bord ne peut recevoir une émission que si elle est suffisamment intense au niveau du satellite. Cela exige l'emploi au sol d'un émetteur puissant, d'une part, et d'une antenne d'émission concentrant l'énergie radio-électrique dans un pinceau étroit, d'autre part.

Cette antenne doit alors être mobile pour que le pinceau puisse suivre le satellite (comme si l'on voulait éclairer en permanence le satellite avec un projecteur situé sur la Terre).

Pour les mêmes raisons, l'émetteur disposé à bord du satellite étant peu puissant, si l'on veut capter correctement l'émission au sol, il est nécessaire de disposer d'une antenne qui ne soit sensible que dans un étroit pinceau; il faut aussi diriger en permanence ce pinceau vers le satellite. (Si l'on utilise une antenne omnidirectionnelle, c'est-à-dire recevant les ondes provenant de toutes les directions, celle-ci recevra non seulement l'émission issue du satellite, mais également tous les parasites extérieurs. L'émission étant peu intense est alors complètement couverte par les parasites, et ne peut être captée même après une forte amplification, car les parasites seraient eux aussi amplifiés. Une antenne très directive, par contre, braquée vers le satellite reçoit la même émission utile, mais très peu de parasites, car seuls ceux qui proviennent de la direction du satellite sont reçus par l'antenne. Il est alors possible de distinguer le signal utile des quelques parasites inévitables reçus également.)

En pratique, la même antenne mobile au sol sert à l'émission et à la réception, mais la nécessité de pouvoir suivre avec précision un satellite qui défile dans le ciel à vitesse assez élevée exige l'emploi d'une installation très importante, comme le montrent les photos de la station de Pleumeur-Bodou, par exemple. (On notera à cet égard que la masse de l'antenne mobile est de l'ordre de 350 tonnes et que l'antenne doit pouvoir tourner à une vitesse maximale de 10 degrés/seconde, et ce avec une précision supérieure au degré.) On conçoit qu'il ne soit guère possible de créer de très nombreuses stations de ce type, ce qui limite le domaine d'emploi de satellites du type de « Telstar » ou de « Relay ».

Les satellites stationnaires, eux, peuvent utiliser des stations au sol plus rudimentaires, car les exigences sur la mobilité de l'antenne sont beaucoup moins sévères.



Le radome de Pleumeur-Bodou.

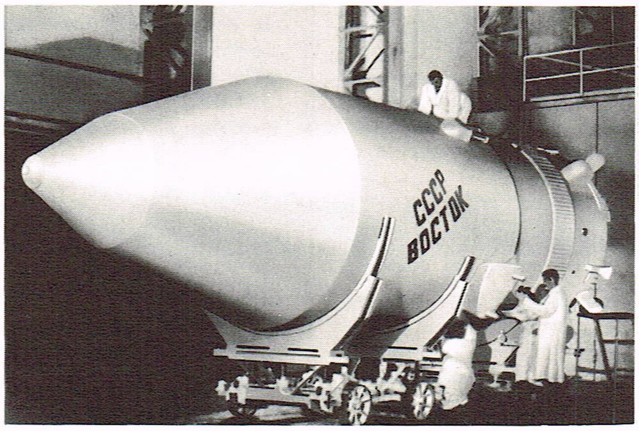
Un satellite de télécommunications du type de « Telstar » peut transmettre un programme de télévision ou environ 300 communications téléphoniques bilatérales; ces performances ne sont rendues possibles qu'au prix d'installations au sol extrêmement importantes.

L'emploi de satellites de plus grande capacité, utilisables avec des stations au sol plus rudimentaires, nécessite que l'on sache augmenter notablement le niveau de puissance électrique disponible à bord du satellite. Au lieu des 30 W (en régime de pointe) disponibles à bord de « Telstar », il serait souhaitable de pouvoir disposer de 500 W environ pour un trafic entre des stations plus rudimentaires que Pleumeur-Bodou par exemple; si, en outre, on souhaite que les émissions transmises par le satellite puissent être reçues directement par les usagers, avec des postes récepteurs du commerce, il est nécessaire de disposer d'une puissance de l'ordre de la dizaine de kilowatts pour des émissions de radio (une centaine de kilowatts pour la télévision).

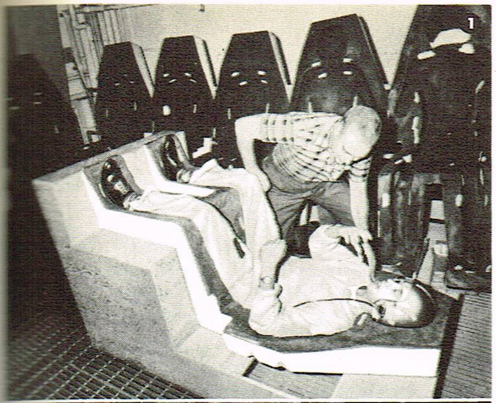
On voit ainsi que la généralisation des télécommunications par satellites sera sans doute liée à la mise au point de sources d'énergie plus puissantes que celles dont on peut disposer à l'heure actuelle sur un satellite.

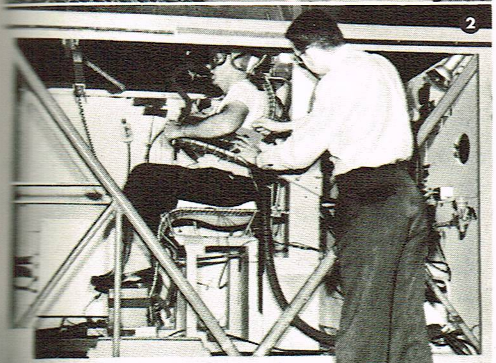


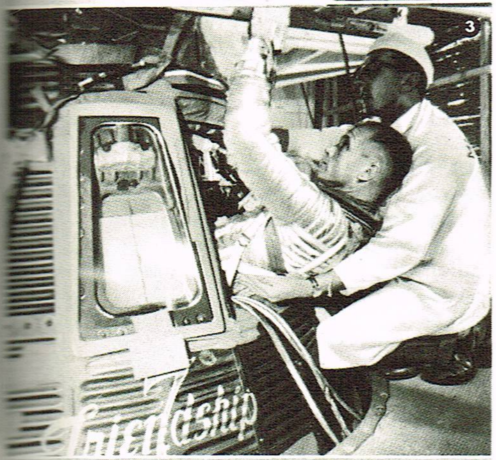
Les six premiers astronautes soviétiques. *De gauche à droite,* **Y. Gagarine, G. Titov, A. Nikolaïev, P. Popovitch, V. Bykovski, V. Térechkova.**



L'engin « Vostok ».







L'opération « Mercury ».

1. **Moulage du corps de A. B. Shepard en vue de la construction du siège de sa capsule ;**

2. **Entraînement de J. H. Glenn à la centrifugeuse ;**

3. **J. H. Glenn prend place dans sa cabine ;**

4. **Lancement de l'astronaute L. G. Cooper ;**

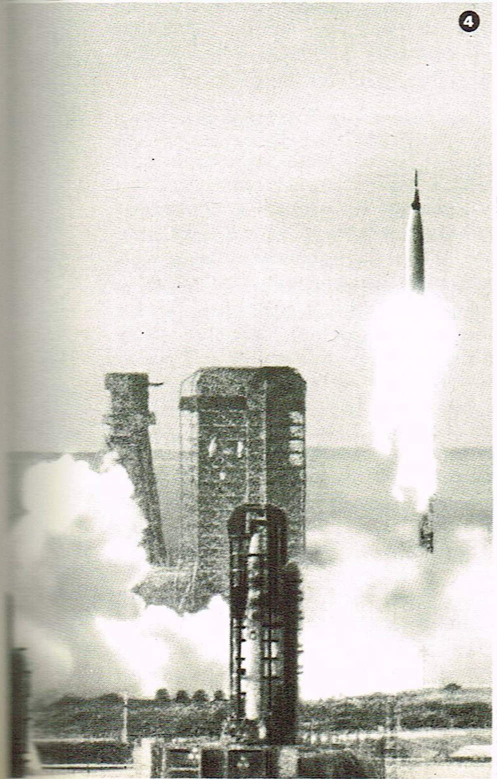
5. **J. H. Glenn dans sa capsule ;**

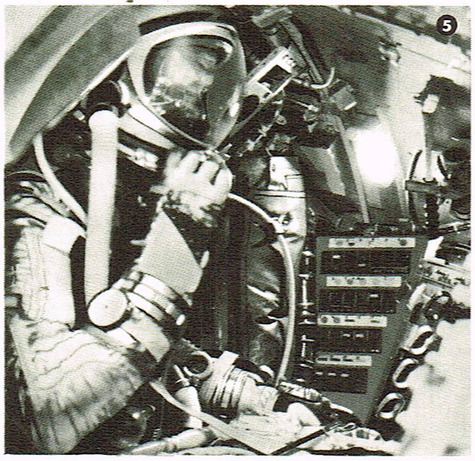
6. **Repêchage en mer de la capsule de Cooper.**

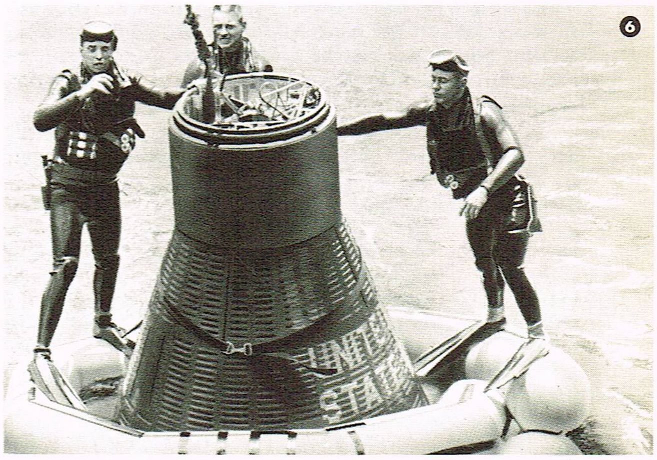
les satellites pilotés

Tous les véhicules (satellites ou sondes spatiales) dont nous avons parlé jusqu'ici étaient des appareils de petite taille et ne contenant que des appareils, parfois fort perfectionnés d'ailleurs. Il est bien évident, cependant, qu'il serait souhaitable que l'homme puisse à son tour entreprendre des missions astronautiques. Avant que de telles missions soient possibles, il est indispensable de pouvoir effectuer de nombreuses expériences permettant d'étudier le comportement d'un être humain dans un véhicule spatial et de définir les conditions d'environnement à créer pour qu'un tel voyage puisse s'accomplir avec le minimum de danger.

Cette nécessité a conduit à la réalisation de satellites scientifiques particuliers à bord desquels des astronautes ont pu effectuer un certain nombre de révolutions autour de la Terre. Ces satellites « Vostok », « Voskhod », « Mercury », « Gemini » ont effectué des missions qui ont permis d'accumuler déjà des renseignements précieux sur le comportement des astronautes et sur les difficultés à surmonter avant de pouvoir effectuer de véritables missions spatiales pilotées. Nous n'entreprendrons pas de donner ici une description détaillée de ces satellites, renvoyant aux illustrations jointes pour l'explication des phases principales de la mission.







Phot. U. S. I. S., Agence de Presse « Novosti ».

comment disposer d'électricité à bord d'un satellite

Tous les satellites dont nous venons de parler contiennent en plus de leur appareillage propre un générateur d'énergie électrique qui, pour les raisons suivantes, doit être considéré comme un élément principal du satellite.

Si le fonctionnement du générateur est interrompu, ou même seulement défectueux, la mission se solde par un échec complet, le satellite devenant totalement inopérant.

Le générateur constitue un élément de poids (et de prix) généralement assez élevé; il est donc souhaitable d'essayer au maximum de le miniaturiser, tout en lui conservant une sûreté de fonctionnement convenable.

Les pages suivantes sont consacrées à l'exposé des problèmes que posent les sources d'énergie à cellules photo-électriques qui, à l'heure actuelle, équipent la quasi-totalité des satellites en service ou en projet.

les cellules photovoltaïques

La photographie de « Tiros I » (p. 129) montre qu'il est presque couvert de cellules photovoltaïques (9 120 cellules exactement). Rappelons-en sommairement le fonctionnement, qui sera étudié plus en détail ultérieurement. De telles cellules sont constituées d'une plaquette de silicium-p (contenant du bore comme impureté) sur laquelle on a fait diffuser du phosphore. Le silicium est ainsi devenu en surface du silicium-n et il existe un champ électrique dans la jonction entre les deux espèces de silicium. Un photon absorbé dans la jonction crée une paire électron-trou et les deux espèces de charges sont triées par le champ électrique de la jonction; les électrons vont dans le silicium-n et les trous dans le silicium-p. Les électrons empruntent alors le circuit extérieur (circuit d'utilisation où ils produisent un courant) pour pouvoir se recombiner avec les trous.

L'emploi d'un tel générateur d'électricité est assez séduisant *a priori;* en réalité, il présente deux défauts importants : un mauvais rendement et une assez grande fragilité. En outre, pour les périodes de passage dans l'ombre de la Terre, il nécessite l'emploi de systèmes de stockage de l'énergie électrique (accumulateurs); ces accumulateurs constituent une part très importante du poids de la source d'énergie et n'ont qu'une sûreté de marche assez faible.

Les causes de mauvais rendement d'une cellule sont généralement les suivantes :
— 6 à 7 p. 100 de la lumière solaire est réfléchie sur la surface de la cellule, et perdue;
— moins de la moitié des photons est absorbée en produisant une paire électron-trou au voisinage de la jonction; les autres photons ne servent qu'à échauffer la cellule;
— une partie des paires créées dans la zone de la jonction se recombinent avant d'être triées, 80 p. 100 seulement des paires étant effectivement triées par la jonction;
— environ 30 à 35 p. 100 de l'énergie électrique ainsi obtenue est effectivement dépensée dans le circuit extérieur, le reste de l'énergie est directement dépensé dans la cellule qu'elle échauffe inutilement.

Finalement, on ne peut espérer de rendement supérieur à

$$0{,}93 \times 0{,}45 \times 0{,}80 \times 0{,}32 = 0{,}11$$

(11 p. 100), et encore ce résultat n'est-il obtenu qu'avec de bonnes cellules de laboratoire. Lorsqu'une fabrication industrielle est entreprise, les cellules fabriquées sont de qualité et de rendement très variables. Il est naturellement possible de trier les cellules pour n'en conserver que les meilleures, mais le prix devient généralement prohibitif (nombre de rebuts trop important) lorsqu'on désire un rendement supérieur à 6 à 8 p. 100.

Puisque la cellule a un faible rendement, elle s'échauffe en cours de fonctionnement. Cet effet est très néfaste, car il diminue la tension aux bornes de la cellule, ce qui a pour résultat de diminuer encore la puissance fournie au circuit d'utilisation. On essaie donc de diminuer cet échauffement en recouvrant la cellule d'un filtre sélectif réfléchissant les rayons lumineux fournissant des paires électron-trou hors de la zone de la jonction (longueur d'onde inférieure à 4 000 Å, ou supérieure à 11 000 Å). On a vu plus haut que ces rayons représentent environ 55 p. 100 de l'énergie solaire incidente.

Lorsqu'une particule de grande énergie (électrons ou protons des ceintures de Van Allen, par exemple) pénètre dans la cellule, elle détruit l'ordonnance du cristal et crée des défauts de structure. Il en résulte une nouvelle baisse du rendement : 25 p. 100 par exemple en trois mois sur un satellite équatorial à 3 000 km d'altitude sous le seul effet des protons. Pour diminuer cet effet, chaque cellule est protégée par une petite plaquette de verre (ou mieux de saphir) destinée à diminuer l'énergie des particules incidentes.

Tous ces dispositifs contribuent naturellement à augmenter le poids et le prix de revient des cellules, de même qu'à diminuer un peu le rendement, une partie de la lumière incidente étant alors absorbée dans les protections.

les batteries d'accumulateurs

A l'altitude à laquelle est situé « Tiros », la proportion maximale du temps passé dans l'ombre de la Terre est de 35 p. 100 environ (aux périodes les plus défavorables de l'année). Puisque la période est voisine de 99 mn, les phases ensoleillées de la trajectoire durent alors 65 mn seulement.

Une batterie d'accumulateurs est chargée par les cellules pendant les phases éclairées de la trajectoire; elle fournit d'une part l'énergie électrique pendant les passages à l'ombre, et d'autre part le supplément de puissance nécessaire pendant les phases éclairées lors de certaines pointes de consommation. Le programme d'utilisation des différents appareils est calculé de telle façon que la batterie soit toujours complètement chargée une fois toutes les vingt-quatre heures (pour éviter un risque de décharge totale). Deux problèmes principaux se posent pour la mise au point de ces batteries d'accumulateurs.

En premier lieu, il faut qu'elles soient parfaitement étanches, sinon l'électrolyte contenu dans la batterie s'évapore et s'échappe dans l'espace, où règne un vide élevé.

On sait, d'autre part, qu'il n'est pas possible de recharger très rapidement une batterie d'accumulateurs (sinon il se produit une électrolyse du liquide et un dégagement gazeux qui, trop intense, provoque la destruction des plaques de l'accumulateur). Or, nous avons vu que sur un satellite comme « Tiros » on ne dispose que de soixante-cinq minutes pour effectuer cette charge entre deux passages à l'ombre. Sur le satellite « Tiros », ainsi, 10 p. 100 seulement de la capacité de la batterie peut être utilisée au cours d'un passage à l'ombre. Les batteries employées sont des batteries étanches nickel-cadmium; leur poids est de 18 kg.

cas d'un satellite " Tiros "

Sur le satellite « Tiros », la surface supérieure couverte de cellules reçoit en moyenne une énergie de 870 W environ (lorsque le Soleil est dans l'axe du satellite); elle restitue une énergie électrique de 35 W, ce qui correspond à un rendement

global de 4 p. 100 seulement (dans les conditions réelles d'utilisation).

La puissance électrique fournie est maximale lorsque la direction du Soleil fait un angle de 45° avec l'axe du satellite (à cause des cellules latérales); elle est alors de 51 W. Le poids total des cellules, connexions et protections diverses est de 11 kg, à l'exclusion de la structure sur laquelle sont fixées les cellules (poids spécifique de 220 kg/kW dans les conditions d'éclairement les plus favorables).

D'une manière générale, on peut considérer que les chiffres indiqués ci-dessus sont valables pour des satellites à basse altitude (bien que la technique ait déjà évolué dans ce domaine depuis le lancement de « Tiros » I). Le poids spécifique d'une telle source d'énergie est voisin de 700 kg/kW, dont 500 kg/kW pour la batterie d'accumulateurs.

Pour des satellites à haute altitude, la durée absolue de la phase éclairée augmente (environ 23 heures pour un satellite stationnaire par exemple). Il est ainsi possible d'augmenter la proportion de décharge de la batterie sur chaque tour (on peut la porter à 60 ou 70 p. 100 par exemple); le poids spécifique de la batterie devient alors voisin de 100 kg/kW. Si l'on tient compte du fait qu'une proportion du temps total plus élevée est passée au soleil, il est également possible de diminuer un peu la puissance des cellules (l'énergie captée pendant la phase éclairée doit être égale à l'énergie totale dépensée par orbite). Le poids spécifique de la source d'énergie est alors voisin de 230 kg/kW (dont 100 kg/kW pour les batteries).

une centrale thermique spatiale ?

On peut constater que dans tous les cas le poids des sources d'énergie précédentes reste assez élevé; ce qui interdit notamment d'envisager l'emploi de systèmes de ce type pour des satellites de puissance élevée (de l'ordre de plusieurs kilowatts par exemple). On notera enfin que l'emploi de cellules photovoltaïques couplées à des batteries d'accumulateurs conduit à des systèmes qui, s'ils sont pratiquement universellement employés actuellement, restent très sensibles aux agents extérieurs; de nombreuses pannes, parfois totales, survenues à bord des satellites lancés à ce jour doivent être imputées à ces sources d'énergie.

Un autre procédé permettant de disposer d'électricité à bord d'un satellite consiste à y installer une véritable centrale thermique composée d'une source de chaleur et d'une machine thermique. Bien que ce système paraisse *a priori* plus complexe que celui des cellules photovoltaïques, il est probable qu'il pourra seul permettre d'obtenir à bord des satellites des puissances électriques suffisamment élevées. C'est à l'examen des problèmes posés par la réalisation d'une telle centrale que sont consacrées les pages suivantes.

utiliser un miroir solaire sur un satellite ?

Le système collecteur d'énergie solaire (ensemble miroir-chaudière) constitue alors simplement une source de chaleur que l'on souhaite maintenir à une température aussi élevée que possible.

Puisque le Soleil est très éloigné de la Terre, un miroir parabolique convenablement orienté peut concentrer les rayons solaires au voisinage de son foyer. Cependant cette concentration ne peut être parfaite pour les deux raisons suivantes :
— le Soleil a un diamètre apparent, et son image ne peut être ponctuelle;
— chaque imperfection du miroir se traduira par une dispersion de l'énergie dans le plan focal, donc par une diminution de la puissance reçue au foyer.

Si nous plaçons une chaudière au foyer du miroir, elle va absorber l'énergie réfléchie par le miroir et s'échauffer. Or, de la même façon que le Soleil, tout corps chauffé rayonne de l'énergie proportionnellement à la quatrième puissance de sa température absolue. Le tableau suivant donne l'ordre de grandeur de cette puissance rayonnée en fonction de la température :

293 °K (température ambiante)	:	0,42	kW/m²
600 °K — —	:	7,6	kW/m²
1 000 °K — —	:	56	kW/m²
6 000 °K (surface du Soleil)	:	73 000	kW/m²

Il est ainsi nécessaire, par exemple, de concentrer l'énergie solaire jusqu'à atteindre un flux de 56 kW/m² (soit environ 40 fois plus que le flux solaire sans miroir) pour porter la chaudière à 1 000 °K. Encore, à cette température et pour ce flux, la chaudière ne pourrait servir de source de chaleur à un moteur, toute l'énergie qu'elle recevrait du Soleil étant perdue à l'extérieur par rayonnement.

Pratiquement, si l'on désire que l'énergie qui est ainsi rayonnée à l'extérieur (et perdue par la chaudière) soit faible comparée à celle qui est absorbée par celle-ci et qui pourra être cédée à une machine thermique, il faut obtenir un facteur de concentration beaucoup plus élevé que celui qui correspond à la condition d'équilibre.

Pour obtenir à cette température un rendement de captation de 90 p. 100 par exemple, il faut disposer d'un miroir concentrant environ 400 fois l'énergie provenant du Soleil.

L'exemple ci-dessus montre les difficultés qu'il faut résoudre pour réaliser un système collecteur (miroir et chaudière) adapté à un véhicule spatial. Ces difficultés sont essentiellement liées à la précision qu'il est nécessaire d'atteindre, dans la construction du miroir et dans son pointage vers le Soleil, si l'on veut obtenir des flux d'énergie importants dans la zone focale du miroir. Cette précision est en général difficilement compatible avec la légèreté que l'on exige en outre de tous les équipements spatiaux.

Enfin il est nécessaire, pendant la phase de lancement de l'engin et en particulier durant la traversée de l'atmosphère terrestre, que le miroir soit protégé par un carénage dont le maître couple soit voisin de celui de la fusée de lancement. On est ainsi obligé, pour les puissances supérieures à quelques centaines de watts, de recourir à des miroirs dépliables. Les problèmes technologiques posés par de tels ensembles sont extrêmement complexes, et aucune solution précise ne s'est encore imposée à ce jour parmi toutes celles auxquelles il est possible de songer : miroirs en pétales, miroirs gonflables, etc.

comment stocker de l'énergie ?

Un véhicule spatial, et en particulier un satellite, peut parfois se trouver « à l'ombre » d'un corps céleste : Terre, Lune, etc.; cela correspond au phénomène, bien connu en astronomie, des éclipses. Un véhicule utilisant l'énergie solaire doit donc être capable de stocker assez d'énergie à bord pour assurer son fonctionnement pendant la durée des passages à l'ombre.

Pour un véhicule à miroir solaire comme celui que nous venons de voir, il est possible de stocker l'énergie sous forme thermique, par exemple en utilisant la chaleur de fusion d'un corps (susceptible d'être très élevée pour certains corps, tels l'hydrure

de lithium LiH, qui fond à 950 °K et restitue en se solidifiant une énergie de 0,7 kWh/kg environ, ou le fluorure de sodium (NaF), qui restitue 0,2 kWh/kg à 1 250 °K, le béryllium : 0,32 kWh/kg à 1 550 °K, et le silicium : 0,4 kWh/kg à 1 700 °K).

Lorsque le stockage s'effectue sous forme thermique, il est naturellement nécessaire d'emmagasiner une quantité d'energie plus élevée (la machine thermique ne transformant cette énergie en électricité qu'avec un mauvais rendement); cependant, un tel système est généralement considéré comme intéressant pour les raisons suivantes :
— la capacité des batteries d'accumulateurs pour des satellites à faible altitude est couramment de l'ordre de 2 à 5 Wh/kg seulement, soit plusieurs dizaines de fois inférieure à celle des batteries à chaleur de fusion;
— grâce au stockage de la chaleur, la machine thermique fonctionne en permanence à un régime constant et uniforme, ce qui est favorable à une grande durée de vie;
— à basse altitude (500 km environ) il faut pendant 65 p. 100 du temps recueillir toute l'énergie nécessaire au fonctionnement pendant un tour complet de la Terre. Le système collecteur doit donc être surdimensionné dans un rapport 1/0,65, soit 1,54, pour tenir compte de la durée des éclipses. Si le stockage se fait sous forme électrique, il faut non seulement surdimensionner le collecteur, mais également toute la machine thermique pour lui permettre durant les phases éclairées de la trajectoire de recharger les batteries d'accumulateurs. Pour une consommation constante de 10 kW par exemple, cela nécessite d'installer un générateur d'une puissance de 15,4 kW (au lieu de 10), qui ne fonctionnera que pendant 65 p. 100 du temps seulement. Il en résultera un nouveau supplément de poids qui défavorisera encore la solution du stockage par batteries d'accumulateurs.

un générateur à radio-isotopes

Les premiers satellites « Transit » étaient alimentés, comme les satellites « Tiros », par un ensemble de cellules photovoltaïques et de batteries d'accumulateurs. Cependant, le succès d'un satellite de navigation à effet Doppler repose essentiellement sur la stabilité de son émetteur, et cette stabilité ne peut être élevée que si l'on dispose d'une source d'énergie elle-même très stable et constante dans le temps. Pour cette raison, il est souhaitable d'utiliser une source d'énergie à fonctionnement permanent, et non cyclique comme toute source solaire (passages à l'ombre de la Terre).

En dehors de l'énergie solaire, on peut utiliser soit l'énergie chimique, soit l'énergie nucléaire; pour des missions de longues durées comme celles de « Transit », il n'est pas possible d'employer des combustibles chimiques. En effet, l'énergie contenue dans l'unité de masse de ces combustibles est faible et il en résulte une trop grande consommation de combustible, d'où un poids élevé. On a donc installé sur les satellites « Transit IV » et suivants un générateur nucléaire à radio-isotopes. On désigne ainsi des atomes dont les noyaux, instables, se décomposent par réaction nucléaire en donnant d'autres noyaux et un rayonnement qui peut être de trois formes :

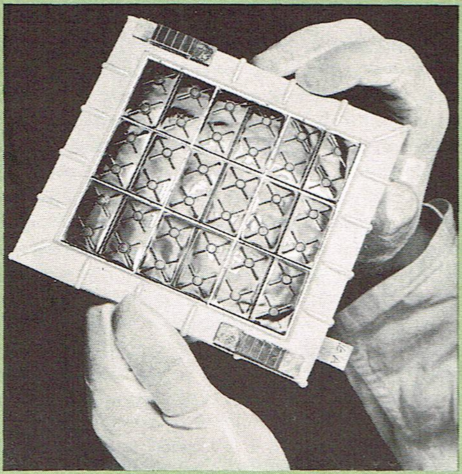
rayons α : noyaux d'atomes d'hélium,
rayons β : électrons,
rayons γ : photons de grande énergie.

Ces rayonnements peuvent être arrêtés dans la masse même de radio-isotopes et dans un écran absorbant, leur énergie étant alors transformée en chaleur. On obtient ainsi une source produisant de la chaleur tant qu'il reste dans le cœur du générateur des noyaux susceptibles de se décomposer. Cette chaleur est ensuite transformée en électricité par une machine thermique.

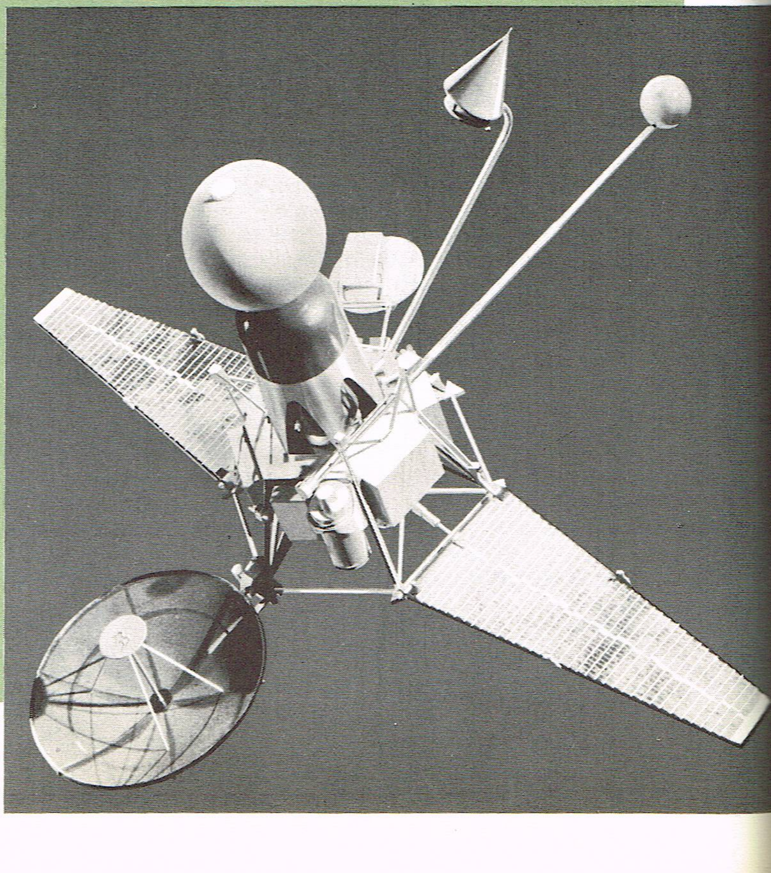
Il est prouvé expérimentalement que pendant un intervalle de temps fixé il se décompose toujours une proportion constante des noyaux instables contenus dans une masse d'un radio-élément. On caractérise donc un radio-élément par sa période, qui est le temps au bout duquel la moitié des noyaux instables présents se seront décomposés. Il en résulte qu'au bout d'un temps égal à la période les rayons émis sont deux fois moins nombreux qu'au début de l'expérience; la puissance thermique du générateur sera donc alors divisée par deux. Comme on souhaite, en général, disposer d'une puissance à peu près constante sur toute la durée d'une mission, il est nécessaire d'utiliser en pratique un radio-élément de période longue comparée à la durée de la mission.

Le satellite « Transit » par exemple, lancé en juin 1961, est prévu

Batteries de cellules photovoltaïques.



Ces deux plans de la sonde « Ranger 3 », malgré leur apparence, ne sont pas des ailes vouées à jouer un rôle aérodynamique, mais des panneaux revêtus de cellules photovoltaïques qui convertissent l'énergie solaire en énergie électrique.



Phot. U. S. I. S.

pour avoir une durée de fonctionnement très grande (une dizaine d'années si possible). Il comporte un générateur à radio-élément dont le « combustible » est du plutonium 238 de période 86,4 ans. La puissance thermique du générateur était de 52 W au moment du lancement; dans dix ans, cette puissance sera encore de 48 W.

Le tableau ci-dessous indique ce que serait devenue cette puissance avec d'autres radio-éléments de période plus faible; il montre l'importance de la période dans le choix d'un radio-élément :

radio-élément	période	puissance au bout de dix ans
Plutonium 238	86,4 ans	48 watts
Strontium 90	28 ans	40,5 —
Curium 244	17,9 ans	35,3 —
Prométhéum 147	2,6 ans	3,58 —
Cérium 144	285 jours	$7{,}3\ 10^{-3}$ —

En plus d'une période convenable, on demandera en général au radio-élément d'avoir une puissance spécifique élevée (puissance thermique par unité de masse de radio-éléments), de ne pas émettre de rayonnements trop difficiles à absorber (rayons γ en particulier, qui nécessitent une importante épaisseur d'écran) et enfin de pouvoir être fabriqué en quantités suffisantes à des prix qui ne soient pas prohibitifs.

A titre indicatif, le tableau ci-dessous donne pour les radio-éléments envisagés à l'heure actuelle la période et la puissance spécifique. On y a également fait figurer des prix approximatifs actuels tels qu'ils ressortent des estimations américaines accessibles dans la littérature (exprimés en francs par watt thermique) :

élément	période	prix (francs/watt)	puissance (watt/gramme)
Cérium 144	285 jours	400	13
Césium 137	33,3 ans	4 500	0,25
Curium 242	162,5 jours	400	122
Plutonium 238	86,4 ans	—	0,6
Polonium 210	138 jours	1 500	140
Prométhéum 147	2,6 ans	15 000	0,3
Strontium 90	28 ans	3 000	0,6

Les radio-éléments constituent une source d'énergie très intéressante pour les applications spatiales; comme semble le vérifier le satellite « Transit », ils permettent notamment une très grande sûreté de fonctionnement sur de longues durées. Leur emploi est essentiellement limité actuellement par les possibilités d'approvisionnement en quantités convenables de corps suffisamment purs.

Ils résultent généralement de l'irradiation par des neutrons de corps non radio-actifs; ils sont donc produits dans des piles atomiques à partir de corps stables introduits à cet effet dans la pile, ou peuvent tout simplement faire partie des résidus de fission qui s'accumulent dans la pile au cours de son fonctionnement.

La quantité de radio-éléments disponibles dans un pays dépend donc essentiellement de la « puissance neutronique » de ce pays (en fait, du nombre et de la puissance de ses piles atomiques) et enfin des moyens de traitement chimique dont dispose ce pays pour séparer les divers radio-éléments produits.

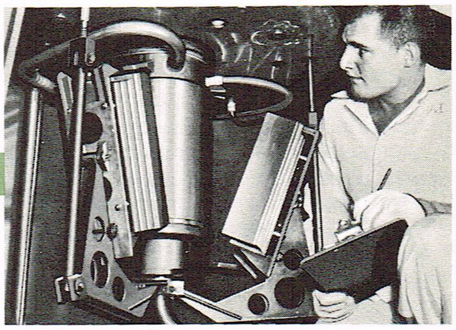
Le coût des investissements correspondants (construction de piles et d'usines de séparation chimique) est trop élevé pour pouvoir être justifié par les seules applications spatiales; **l'emploi de générateurs à radio-éléments est ainsi lié au stade de développement de l'industrie atomique, que ce soit en moyens de production ou en débouchés autres que ceux de l'industrie spatiale.**

vers le réacteur nucléaire spatial

On peut se débarrasser des sujétions économiques précédentes en installant directement à bord du véhicule spatial un réacteur (pile) atomique. Malheureusement, une telle solution ne peut être intéressante que lorsqu'on désire obtenir des puissances thermiques élevées. La taille minimale d'un réacteur nucléaire à fission est, en effet, imposée par les « conditions critiques » (pour le maintien d'une réaction en chaîne dans le réacteur). **Il existera donc toujours un poids minimal de réacteur nucléaire,** que l'on ne peut guère diminuer, même lorsqu'on souhaite obtenir une puissance faible. Ce phénomène est mis en évidence dans le tableau suivant donnant les poids du réacteur pour trois prototypes américains en cours de réalisation, en fonction de la puissance thermique de ces réacteurs. Ces poids correspondent au réacteur nu, avec réflecteur, mais sans écran de protection contre les radiations.

« SNAP 10 A »	30 kW thermiques	115 kg
« SNAP 2 »	50 kW —	115 kg
« SNAP 8 »	600 kW —	185 kg

SNAP 2



Il est possible d'indiquer dès maintenant quelques caractéristiques principales des futurs réacteurs nucléaires spatiaux :

1. On verra plus loin qu'il est important que le réacteur soit capable de fonctionner à la température la plus élevée possible. Cela interdit d'utiliser l'eau lourde comme modérateur (elle devrait être soumise à une pression considérable pour ne pas bouillir); les réacteurs américains indiqués plus haut utilisent de l'hydrure de zirconium comme modérateur, les points les plus chauds du réacteur étant portés à une température voisine de 550 °C pour le « SNAP 10 A », 700 °C pour le « SNAP 2 » et 760 °C pour le « SNAP 8 ». Pour les températures plus élevées, il faudrait alors utiliser des réacteurs modérés au béryllium ou des réacteurs rapides à carbure d'uranium.

2. Il est naturellement souhaitable que le réacteur soit le plus léger possible. Ce résultat est atteint en diminuant la masse critique par l'emploi de réflecteurs de neutrons entourant le cœur de la pile. Ces réflecteurs sont en béryllium sur les réacteurs « SNAP » et peuvent constituer une partie importante du réacteur comme le montrent les chiffres suivants relatifs au réacteur « SNAP 8 » :

Longueur du cœur (cylindrique)	: 48 cm
Diamètre du cœur	: 38 cm
Epaisseur du réflecteur	: 8 cm.

Toujours pour diminuer le poids du réacteur, il est souhaitable de simplifier au maximum les dispositifs de contrôle. Sur les réacteurs à réflecteur, ce contrôle peut être fait simplement (et dans une zone peu chaude du réacteur) en agissant directement sur les réflecteurs, dont une partie est mobile et permet de régler une « fuite de neutrons » vers l'extérieur.

3. Le troisième point important est celui de la protection contre les rayonnements émis par le réacteur. Tant que le réacteur n'a pas fonctionné, il n'émet pratiquement pas de rayonnements;

Phot. U. S. I. S.

ceux-ci se produisent, par contre, en cours de fonctionnement et après l'arrêt du réacteur (rayonnement des résidus de fission). Les problèmes de protection sont de deux sortes :
— protéger le véhicule du rayonnement durant la mission (pendant le fonctionnement du réacteur);
— si le véhicule revient sur Terre, éviter qu'il y ait contamination de la Terre par les produits de fission.

a) Pendant la période de fonctionnement, le réacteur est dans le vide et les rayonnements se propagent en ligne droite (pas de dispersion par l'atmosphère). On peut donc protéger tout le véhicule en le plaçant dans le cône d'ombre d'un écran partiel situé entre le réacteur et lui. Le poids de l'écran dépend alors de la distance entre le réacteur et la zone à protéger, et en outre de la sensibilité au rayonnement des éléments à protéger. Pour le réacteur « SNAP 8 », par exemple, en supposant que le réacteur est à 6 m de la zone à protéger, il faut au minimum un écran (hydrure de lithium) pesant une tonne pour un être humain et 150 kg pour des appareils électroniques transistorisés courants; ces poids deviendraient respectivement 850 kg et 120 kg pour une distance de 9 m.

b) La seule méthode envisagée actuellement pour se protéger des radiations au retour du réacteur sur la Terre est de détruire le réacteur par échauffement thermique au cours de la rentrée dans l'atmosphère. Tous les matériaux constituant le réacteur, ainsi que les produits de fission, sont alors répandus et dilués dans les couches élevées de l'atmosphère (altitudes supérieures à 60 km), où ils créeraient une activité négligeable devant celle qui existe naturellement ou par suite des essais nucléaires. Si ce problème n'est pas essentiel dans les années à venir, où un nombre faible de réacteurs sera ainsi utilisé, d'autres méthodes devront sans doute être envisagées à plus longue échéance en raison de l'augmentation du nombre et de la puissance des réacteurs (il peut, en outre, se révéler difficile de détruire de façon convenable des réacteurs de masse importante par ce procédé).

une machine à vapeur pour satellites

Lorsqu'on utilise un réacteur nucléaire ou un four solaire à miroir, on dispose à bord du véhicule d'une source de chaleur à partir de laquelle on souhaite obtenir de l'énergie électrique. Cela sera réalisé au moyen d'une machine thermique dont nous indiquerons d'abord les conditions générales de fonctionnement. En vertu du principe de Carnot (second principe de la thermodynamique), une machine thermique emprunte de la chaleur à une source chaude, en transforme une certaine partie en travail et restitue l'énergie non transformée à une source froide. La proportion maximale d'énergie prise à la source chaude qu'il est possible de transformer en travail est égale à $R = 1 - T_0/T_1$, si T_1 est la température (absolue) à laquelle l'énergie est prise à la source chaude et T_0 celle à laquelle elle est restituée à la source froide.

Sur Terre, la température minimale de source froide est la température ambiante (soit celle de l'air, soit celle d'une rivière, par exemple). Dans l'espace, par contre, cette température peut être rendue aussi faible que l'on veut. Puisqu'il n'y a pas d'air dans l'espace, les échanges de chaleur ne peuvent se faire que par rayonnement thermique. En un point de l'espace, par exemple, un corps reçoit le rayonnement du Soleil, le rayonnement thermique de la Terre, ainsi que la partie du rayonnement solaire diffusée par l'atmosphère terrestre, enfin le rayonnement de tous les corps célestes. Si l'on excepte le rayonnement provenant du Soleil et celui qui est diffusé par l'atmosphère terrestre, le rayonnement reçu par un corps est, en fait, pratiquement négligeable.

Cela veut dire que si l'on place une plaque plane, par exemple, à l'ombre d'écrans la protégeant du rayonnement solaire et de celui qui est diffusé par l'atmosphère terrestre, quelle que soit la température de cette plaque plane, elle rayonnera dans l'espace plus d'énergie qu'elle n'en recevra des autres corps célestes. On voit ainsi que dans l'espace il est pratiquement possible de céder de l'énergie à l'extérieur, à une température aussi faible qu'on veut, et cela de façon illimitée.

En fait, il n'est malheureusement pas possible d'utiliser cet avantage pour augmenter le rendement de la machine thermique. Nous avons vu, en effet, en parlant des collecteurs solaires, qu'un corps rayonnait une énergie proportionnelle à sa surface et à la quatrième puissance de sa température absolue. Imaginons, par exemple, une machine dont le radiateur soit à une température de 10 °K; la puissance rayonnée est alors de $5{,}67\ 10^{-4}\ W/m^2$. Si l'on veut que la machine fournisse une puissance de 1 kW avec un rendement de 98 p. 100 (source chaude à 500°), elle doit rejeter 21 W au radiateur à 10 °K. La surface nécessaire est alors de 37 000 m² pour 1 kW utile produit; cette valeur de la surface est absolument prohibitive, et il en sera de même du poids du radiateur correspondant.

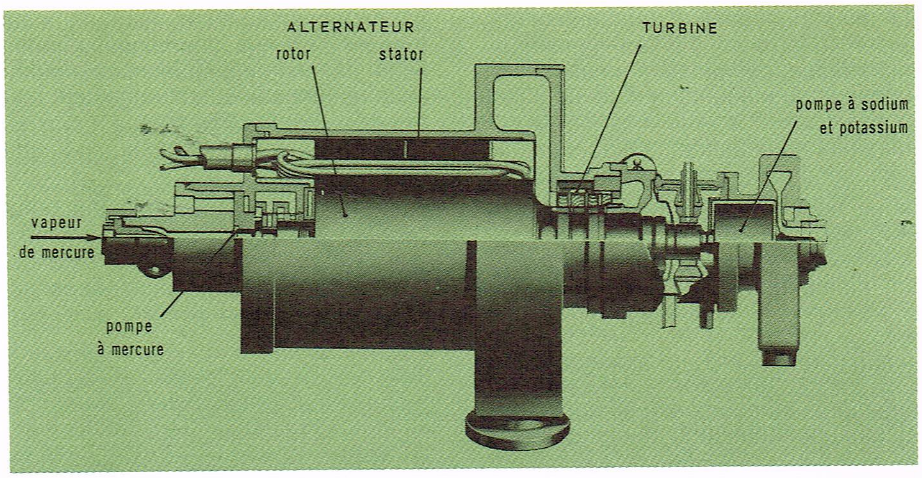
Il ne faut pas, pour des applications spatiales où le poids constitue le paramètre essentiel, accorder une importance trop fondamentale à la notion de rendement. Dans le cas d'un réacteur nucléaire de faible puissance, par exemple, le poids du réacteur est presque indépendant de la puissance; il est voisin du poids minimal représentant la taille critique du réacteur. La notion de rendement n'a plus d'intérêt dans ce cas, et il vaut beaucoup mieux tenter de diminuer la surface, et donc le poids du radiateur, qu'augmenter le rendement.

Ce poids minimal de radiateur est obtenu lorsqu'on prend un rendement de 25 p. 100 seulement. Dans le cas de l'exemple précédent, la température de radiateur deviendrait 375 °K. Pour obtenir 1 kW, il faut en consommer 4 à la source chaude, et en évacuer 3 au radiateur, ce qui exige (à 375 °K) une surface de 2,7 m² (au lieu de 37 000 m² précédemment).

Si, par contre, la source chaude était à 1 000 °K (au lieu de 500), un rendement de 25 p. 100 serait atteint pour une température de radiateur de 750 °K, et pour obtenir une puissance de 1 kW il suffirait d'une surface de radiateur de 0,17 m². Ces quelques exemples numériques simples montrent pourquoi, en pratique, la température de radiateur sera toujours choisie assez élevée, alors que les conditions régnant dans l'espace permettraient d'atteindre théoriquement des rendements très élevés grâce à des températures de radiateur très faibles.

Dans le cas des générateurs à collecteur solaire, il est naturellement nécessaire d'augmenter la surface de miroir (et la quantité d'énergie à stocker pour les durées d'éclipse) lorsque le rendement de la machine diminue. La température de radiateur correspondant à un poids de générateur minimal est alors plus faible que celle que nous avons indiquée plus haut pour les réacteurs nucléaires; elle est assez variable selon les procédés technologiques utilisés, mais ne descend que rarement au-dessous de la température fournissant un rendement de 40 p. 100.

Les machines thermiques utilisées peuvent reproduire à petite échelle les installations existant dans les centrales électriques thermiques employées sur terre. Les seules différences proviennent des faits suivants :
— la « source froide » du cycle thermique est à température assez élevée (au moins 600 à 700 °K suivant les sources de chaleur utilisées);
— l'obtention d'un poids minimal est généralement recherchée, fût-ce au prix d'une diminution du rendement;
— les deux conditions précédentes imposent que toute la machine fonctionne dans une zone de température beaucoup plus élevée que celle qui correspond aux machines terrestres. Il n'est alors pas pratique d'utiliser l'eau et sa vapeur comme fluide circulant dans une telle machine (ce qui correspondrait à des pressions de fonctionnement très élevées); aussi l'emploi de métaux liquides s'est-il généralisé : mercure, rubidium, sodium, etc. Un système de conversion fonctionnant d'après ce principe a été représenté à la page ci-contre.



La turbine produisant l'énergie mécanique issue du cycle thermique entraîne un alternateur; ce dernier fournit l'énergie électrique désirée.

comment supprimer turbines et alternateurs

L'emploi de cycles à vapeur, conjugués avec une turbine et à un alternateur, représente une solution classique qui est la transposition de celle que nous utilisons sur terre. Cette formule conduit soit à des réalisations lourdes, soit à des éléments fragiles et très complexes lorsqu'on essaie de diminuer les poids.

D'autres solutions moins détournées sont actuellement à l'étude pour produire de l'électricité à partir d'une source de chaleur. Elles consistent toutes à essayer de se débarrasser de l'intermédiaire mécanique représenté par l'ensemble turbo-alternateur.

De telles solutions étaient déjà étudiées pour la production d'énergie électrique terrestre, mais l'intérêt de leur emploi pour des applications spatiales (légèreté et plus grande sûreté de marche obtenue par la suppression de toutes les pièces mécaniques mobiles) a donné une nouvelle impulsion aux recherches, qui se poursuivent activement dans de nombreux pays.

Supposons un circuit électrique comportant deux rails et un conducteur AB pouvant glisser sur les rails; un champ magnétique perpendiculaire au plan du circuit est en outre appliqué. Si l'on déplace le conducteur AB parallèlement aux flèches ci-dessous, un courant circule dans le circuit électrique; ce courant est dû au fait que le mouvement de la barre AB oblige les particules chargées (électrons) contenues dans la barre à se déplacer dans le champ magnétique.

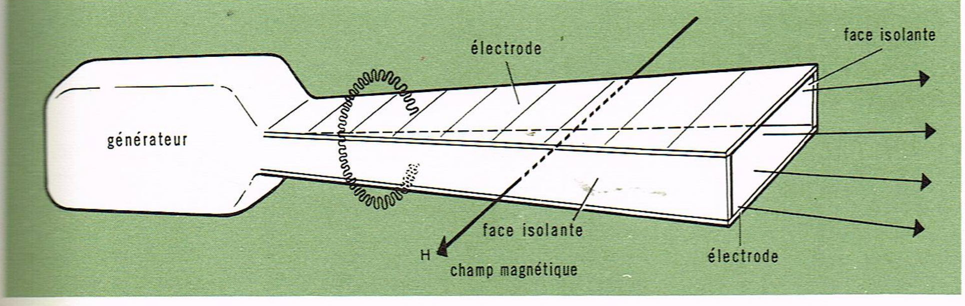
Imaginons alors qu'au lieu de déplacer la barre AB avec la main on la tire avec un treuil actionné au moyen d'une turbine alimentée par un générateur de vapeur, nous avons réalisé de manière rudimentaire un turbo-alternateur comportant :
— une source de gaz sous pression;
— une tuyère accélérant les gaz (énergie thermique → énergie mécanique);
— une turbine recueillant l'énergie cinétique des gaz;
— un alternateur où l'énergie mécanique de la turbine est utilisée à déplacer un conducteur dans un champ magnétique (énergie mécanique → énergie électrique).

Nous allons modifier les éléments du circuit tout en conservant le même principe de base qui consiste à obliger des charges à se déplacer dans un champ magnétique.

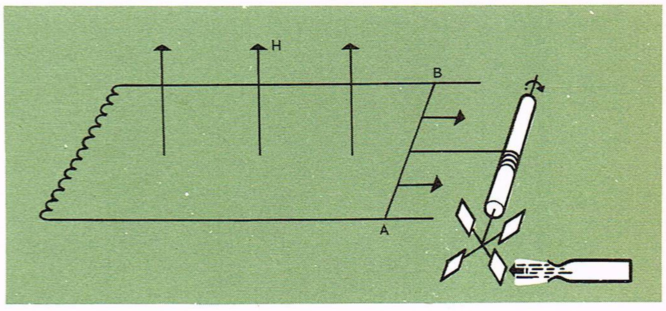
Supposons que nous supprimions la barre AB et que nous la remplacions par un tube isolant contenant un gaz chaud ionisé; le tube peut glisser entre deux plaques conductrices qui ferment les deux bouts du tube et remplacent les rails précédents. Un gaz ionisé est un mélange de molécules et d'atomes qui par suite de l'agitation thermique ont perdu un ou plusieurs électrons périphériques, devenant ainsi des atomes chargés positivement, ou ions positifs; le gaz contient enfin les électrons qui ont été arrachés aux atomes; ces trois espèces de corps sont supposés en équilibre thermique, c'est-à-dire que l'énergie d'agitation thermique est telle qu'il se crée en moyenne par seconde autant de nouveaux ions qu'il s'en détruit d'anciens (par attraction électrostatique entre les ions et les électrons et formation d'un atome neutre ordinaire).

Comme précédemment, on ferme le circuit par un conducteur et on applique un champ magnétique. Si l'on déplace le tube, il va de nouveau circuler un courant électrique, le cylindre de gaz ionisé jouant le rôle du conducteur AB, et les ions et les électrons contenus dans le tube étant obligés de se déplacer dans le champ magnétique par suite du mouvement du tube.

Générateur magnétohydrodynamique.



Barre sur deux rails (schéma de principe).



On voit ainsi qu'il suffit d'obliger un gaz conducteur (ionisé) à circuler entre deux électrodes dans un champ magnétique pour produire un courant électrique; cela est obtenu en prenant un générateur de gaz et une tuyère comprenant deux faces conductrices et deux faces isolantes; un champ magnétique est appliqué perpendiculairement aux faces isolantes.

Si l'on réunit les deux électrodes par un circuit, il passe du courant dans ce circuit lorsqu'on met le générateur en fonctionnement. Il suffit de considérer par la pensée de très nombreux tubes de gaz allant d'une électrode à l'autre et obligés par l'écoulement aérodynamique dans la tuyère de se déplacer dans le champ magnétique, pour se rendre compte que l'on a ainsi reproduit de façon élégante le montage de la première expérience. L'intérêt de cette solution est qu'elle supprime la turbine, le treuil et la barre AB, c'est-à-dire en définitive toutes les pièces mécaniques mobiles du turbo-alternateur. Le générateur réalisé est ce que l'on appelle un générateur magnétohydrodynamique; aucun appareil de ce type n'est malheureusement encore parvenu à ce jour au stade des applications pratiques.

les convertisseurs statiques

Il est même possible de supprimer également la vapeur circulant dans le convertisseur magnétohydrodynamique et de réaliser une machine thermique entièrement statique. Sur « Transit IV » on a même utilisé un tel convertisseur, dit « convertisseur thermo-électrique »; cela a permis, en liaison avec une source de chaleur à radio-isotopes, de réaliser un générateur d'électricité très stable, de grande durée de vie et parfaitement adapté à l'alimentation d'un émetteur de radio à fréquence très constante. Le fonctionnement de tels convertisseurs peut se décrire comme suit : on sait qu'un corps solide conducteur de l'électricité peut être grossièrement représenté comme un ensemble d'électrons susceptibles de se déplacer assez librement à l'intérieur d'un réseau composé d'ions positifs. Un électron est attiré par des ions positifs, mais à l'intérieur du conducteur un électron étant entouré de tous côtés par des ions positifs, la résultante de toutes ces attractions sera pratiquement nulle en moyenne; cela explique que les électrons puissent circuler librement dans le cristal. Au voisinage du bord du cristal, par contre, tous les ions se trouvent du même côté de l'électron (vers l'intérieur du cristal), la résultante des forces d'attraction ramène donc l'électron vers l'intérieur du cristal; cela explique que les électrons, bien que circulant facilement dans le cristal, ne puissent s'en échapper que s'ils ont une très grande énergie.

Supposons que l'on porte un cristal à haute température; on va augmenter l'énergie d'agitation des électrons et certains arriveront à échapper à l'attraction des ions, ils seront émis par le cristal. Naturellement, pour chaque électron ainsi émis, la différence entre le nombre d'ions et d'électrons dans le cristal augmentera d'une unité; le cristal prendra une charge globale positive et retiendra de plus en plus solidement les électrons intérieurs. Pour une température donnée, il existe ainsi un état d'équilibre correspondant à un cristal ayant perdu un nombre donné d'électrons.

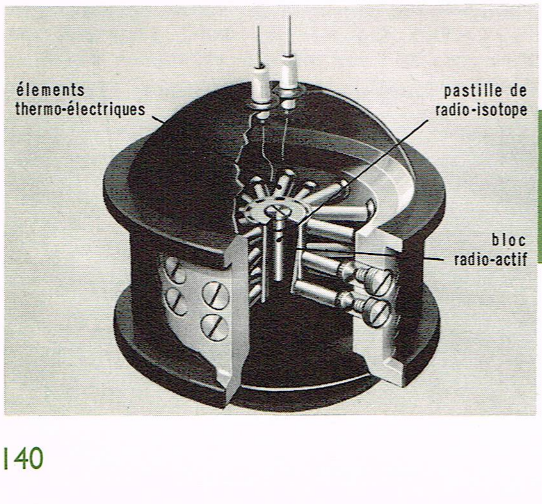
Au lieu de laisser ce cristal chaud complètement isolé, supposons que l'on recommence l'expérience en mettant au voisinage l'un de l'autre deux cristaux identiques, dont l'un est chauffé et l'autre refroidi. Les deux cristaux seront supposés être dans le vide et l'on s'est assuré avant de commencer le chauffage que les deux cristaux sont bien au même potentiel électrique (en les mettant en contact pendant quelques instants, par exemple).

Lorsque le cristal chauffé est à une température suffisante, il émet un électron comme précédemment; cet électron qui se trouve dans le vide va être attiré par les ions du cristal froid et va pénétrer dans ce cristal. Le cristal chaud contiendra globalement une charge positive supplémentaire et le cristal froid une charge négative; cela veut dire qu'il va s'exercer maintenant un champ électrique entre les deux cristaux. Pour qu'un deuxième électron puisse passer du cristal chaud au cristal froid, il faut qu'en plus de l'attraction qu'avait dû vaincre le premier électron il surmonte l'attraction de l'ion positif supplémentaire existant dans le cristal chaud et la répulsion de l'électron supplémentaire du cristal froid. Pour un couple de températures donné des deux cristaux, il y a ainsi un état d'équilibre, le cristal chaud possédant *n* ions supplémentaires et le cristal froid *n* électrons. Les deux cristaux ne sont alors plus au même potentiel électrique et si on les relie tous deux par un circuit électrique, ce circuit sera parcouru par un courant, les électrons quittant le cristal froid pour rejoindre les ions du cristal chaud. On constate que l'on a réalisé ainsi une source de courant.

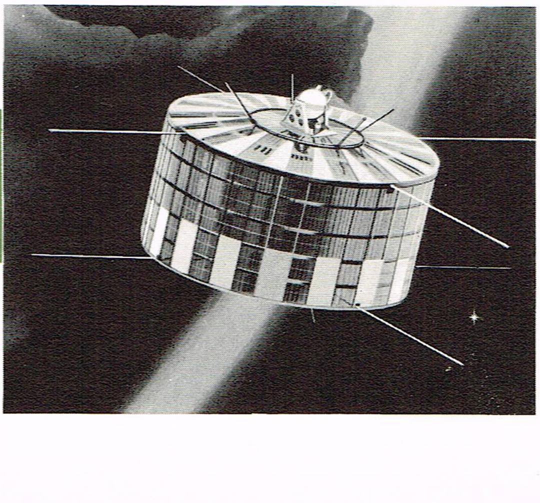
Il est essentiel de remarquer que pour que ce dispositif fonctionne il faut satisfaire aux deux conditions suivantes :
a) apporter assez d'énergie thermique au cristal chaud pour qu'il puisse émettre des électrons;
b) refroidir suffisamment le second cristal pour empêcher les électrons qui s'y trouvent de pouvoir en sortir (il faut naturellement permettre le passage des électrons du cristal chaud au cristal froid, mais empêcher le mouvement inverse).

On a ainsi réalisé une machine thermique particulière dont le fluide moteur est le « gaz d'électrons » contenu dans un conducteur; ce gaz gagne de la chaleur dans le cristal chaud (chaudière), fournit de l'énergie électrique en passant d'un cristal à l'autre et perd sa chaleur dans le cristal froid (radiateur).

Cette machine thermique spéciale est un convertisseur thermoionique. Pour qu'un tel convertisseur fonctionne, il faut que le cristal chaud soit à très haute température (1 500 °C, par exemple). On peut, par contre, remplacer le vide séparant les deux cristaux par un cristal unique d'un matériau conducteur différent et soudé à ses deux extrémités aux cristaux précédents.



Coupe du générateur SNAP 3. La photographie montre cette source d'énergie électrique installée dans le satellite « Transit IV ».



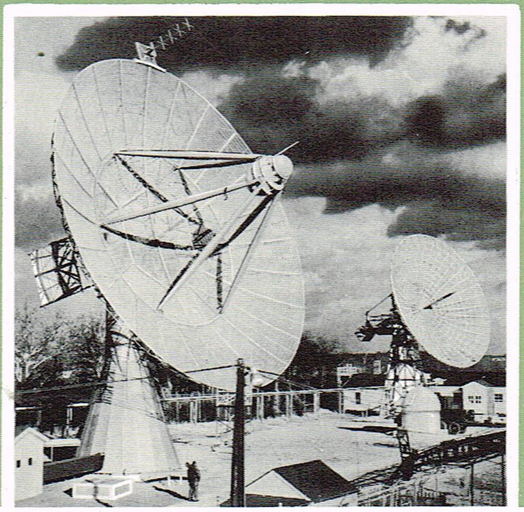
Le système obtenu est rigoureusement analogue, mais il suffit maintenant d'une température plus faible pour qu'un électron passe du cristal chaud dans le cristal intermédiaire (en effet, les ions du cristal intermédiaire attirent un électron se trouvant dans le cristal chaud à la frontière; l'électron n'a plus à surmonter que la différence des deux attractions pour passer dans le cristal intermédiaire).
Un convertisseur thermo-ionique dans lequel le vide a été remplacé par un conducteur de nature différente s'appelle un « convertisseur thermo-électrique ». Il peut produire un courant avec une différence de température plus faible qu'un convertisseur thermo-ionique. La mise au point de ces générateurs est activement poussée à l'heure actuelle et il est probable qu'ils trouveront dans l'avenir de nombreuses applications spatiales; à cet égard, le convertisseur thermo-ionique surtout peut se révéler intéressant, car il est bien adapté au fonctionnement à température très élevée, ce qui, comme nous l'avons vu, est utile pour diminuer les poids des sources d'énergie.

Au cours des pages précédentes nous avons abordé deux aspects particuliers des réalisations spatiales : celui des véhicules de lancement et celui des satellites et des sondes spatiales. Encore sur chacun de ces points nous a-t-il été nécessaire de nous restreindre à deux techniques particulières : celle de la propulsion d'abord, puis celle de la génération d'énergie de bord. Comme nous l'avons signalé dans l'introduction, il aurait été possible de mettre en lumière autant de problèmes sur un grand nombre d'autres techniques, comme l'électronique, le guidage, la construction de la structure elle-même des véhicules, etc. Ce qu'il est convenu d'appeler la technique spatiale n'est, en effet, que la juxtaposition et la synthèse d'un grand nombre de techniques très diverses, la caractéristique commune de toutes ces contributions étant qu'elles représentent, chacune dans leur domaine propre, l'état de développement le plus avancé compatible avec une sûreté de fonctionnement convenable.

Cette recherche des performances les plus élevées possibles ne correspond naturellement pas à un simple désir de perfection, mais est absolument nécessaire à l'accomplissement des missions spatiales : un satellite comme « Telstar » par exemple ne pesait que 77 kg et pourtant il contenait un appareillage rendant les mêmes services qu'une station-relais terrestre, qui, elle, pèserait plusieurs tonnes. A titre indicatif, et pour donner une idée de la complexité de tels satellites, on notera que « Telstar » comportait environ 2 500 diodes et transistors, 1 200 condensateurs, 2 800 résistances, 340 inductances, 70 transformateurs, 10 relais, etc.

Lorsqu'on tient compte de cette complexité, il n'est alors plus étonnant de voir l'ampleur des travaux de tous ordres nécessaires à la réussite d'une mission. En dehors des études ou essais de mise au point des composants eux-mêmes, il est, en effet, nécessaire de procéder à de multiples essais sur l'ensemble du véhicule avant son lancement. Pour des missions compliquées, ces essais au sol peuvent éventuellement durer plusieurs mois et immobiliser une aire de lancement pendant des semaines.

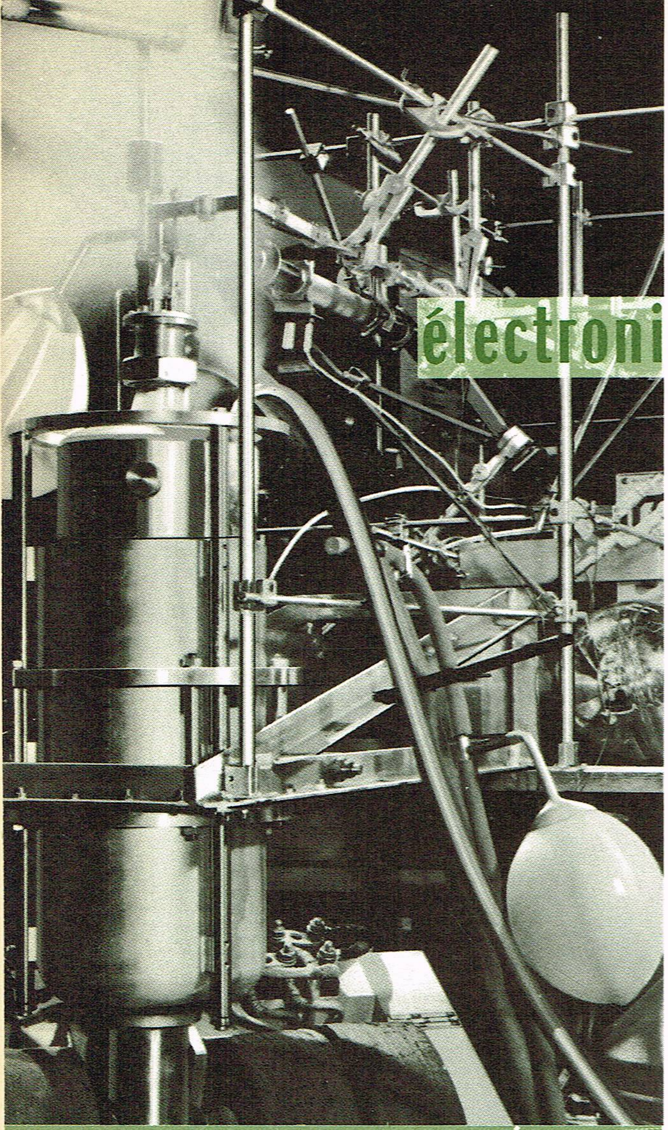
Enfin, on a coutume de s'émerveiller des performances des véhicules lanceurs ou des satellites, mais il est indispensable de se persuader que le succès des missions dépend dans une large mesure d'une infrastructure considérable. Celle-ci a dû également être bâtie de toutes pièces aux seules fins de la recherche spatiale. Une installation comme la station de Pleumeur-Bodou, en Bretagne, constitue un exemple typique des investissements qu'il est nécessaire de consentir pour permettre les expériences spatiales. Le lecteur a pu, en parcourant les illustrations et les tableaux de ce chapitre, se convaincre de l'importance des moyens nécessaires au succès d'une mission du type de celle de la capsule « Mercury », par exemple. Nous souhaitons qu'ils montrent que la technique spatiale n'est pas un domaine d'activité isolé, mais bien, au contraire, le point de rencontre de toutes les techniques les plus diverses et le reflet du niveau de perfectionnement et de développement de toute l'industrie.



Station installée dans le New Jersey pour le repérage et la poursuite des satellites « Tiros ».

Phot. U. S. I. S.

ÉLECTRONIQUE



Le premier maser expérimental conçu en France.

électronique

Mot à la mode. Au point de servir de thème à maints slogans commerciaux.

Mais, beaucoup plus profondément, mot clé de la révolution que vit notre génération.

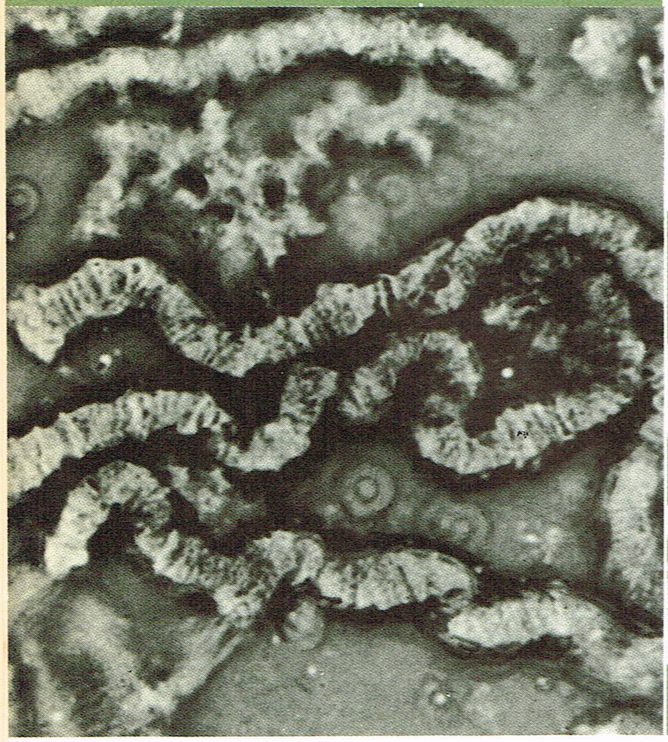
Le XIXe siècle avait vu un premier et important bouleversement s'opérer dans la condition humaine : la découverte des richesses du sous-sol, et principalement du charbon, inaugurait l'ère industrielle. Dès lors commençait une course de plus en plus effrénée vers une production maximale d'énergie. Le niveau de vie d'un peuple se chiffrait en kilowatts-heures; l'homme domestiquait des quantités croissantes d'énergie, doublant sa consommation tous les dix ans, la décuplant à chaque génération. Et au moment même où se dessinait, dans un horizon plus ou moins lointain, le spectre d'un épuisement définitif des ressources énergétiques exploitables de la planète, la relève était assurée par l'énergie nucléaire et par l'énergie solaire, toutes deux inépuisables et bientôt domestiquées. L'homme était devenu le maître d'une puissance pratiquement illimitée.

Mais que sert d'accroître indéfiniment la puissance mise à notre disposition si l'on ne peut ni la commander ni la diriger? Que sert de savoir, si l'on ne peut transmettre ses connaissances? L'homme, non content de ce que sa supériorité sur les autres êtres vivants se soit traduite par l'acquisition d'un langage, désire communiquer avec ses frères, de plus en plus, de plus en plus vite, de plus en plus loin. A la faim d'énergie qui caractérisait le siècle dernier succédait une faim plus impérieuse encore d'information.

Le modeste montage à galène cher aux premiers amateurs de T. S. F. cédait bientôt la place à la lampe diode, puis à la triode. Mais il fallut encore attendre un demi-siècle pour qu'éclatât la révolution qu'apportèrent les composants à semi-conduction et que s'accélérât le progrès technologique : quinze ans exactement après l'invention du transistor (1948), l'étroit faisceau d'un laser permettait à « Telstar » de retransmettre le premier spectacle de Mondovision, et l'on pouvait envisager le transport simultané de 100 000 communications téléphoniques, ou de plusieurs milliers d'émissions télévisées. Quel progrès sur les pesants et coûteux câbles transatlantiques!

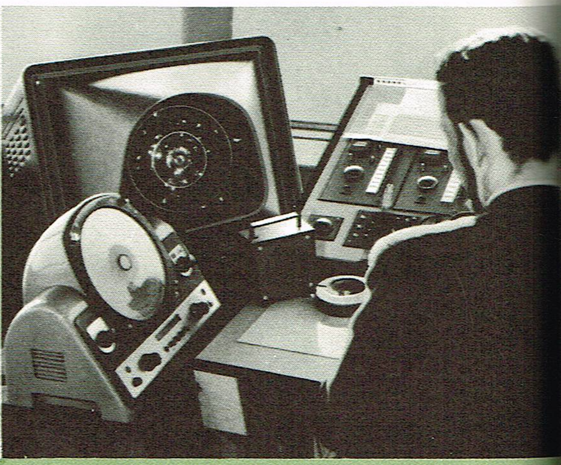
Pratiquement, une liaison instantanée peut être établie entre tout point du globe : il suffit d'installer émetteur et récepteur, la liaison étant réalisée par quelque satellite de télécommunication. Ainsi se trouve réalisée la prophétie de Paul Valéry : l'humanité vit dans un monde fini et en état de totale intercommunication, sinon de totale intercommunion. Ainsi est résolu le problème des télécommunications (aux incidences budgétaires ou financières près, ainsi que le savent les Français, candidats par centaines de milliers

Chromosomes des glandes salivaires vus en contraste de phase.





Transistors.



Appareillage électronique à la tour de contrôle d'O-

Phot. C. S. F., René Bouillot.

et INFORMATION

au téléphone). Problème qui touche de si près l'activité politique que le ministère des Postes, Télégraphes et Téléphones renonce à son titre traditionnel — mais non au sigle le désignant encore — pour devenir, sous le nom de « ministère des Postes et Télécommunications », un rouage essentiel de gouvernement, et que le ministère de l'Information voit son rôle grandir sans cesse. N'est-ce pas également sous cette rubrique que doivent se ranger les accords les plus étroits de notre coopération avec les nations africaines?

Transistors et lasers sont donc à la base de la véritable mutation qui s'opère dans la civilisation : seuls les progrès qu'ils permettent rendent possible la réalisation d'une véritable communauté mondiale. Ce n'est pas sans mal que l'homme a appris, siècle après siècle, à vivre la vie communautaire du clan familial, puis — successivement — celle de la cité antique, de la province, puis de la nation. Le passage de cultures étroitement et exclusivement nationalistes à une vision planétaire réellement vécue, où tout événement acquiert une résonance planétaire immédiate, et où nul ne peut prétendre ignorer la vie de ses frères, fussent-ils distants de milliers de kilomètres, risque d'être encore long et laborieux. Mais les progrès de l'électronique, permettant la transmission de plus en plus rapide et complète de l'information, conduira nécessairement l'humanité à cette profonde mutation.

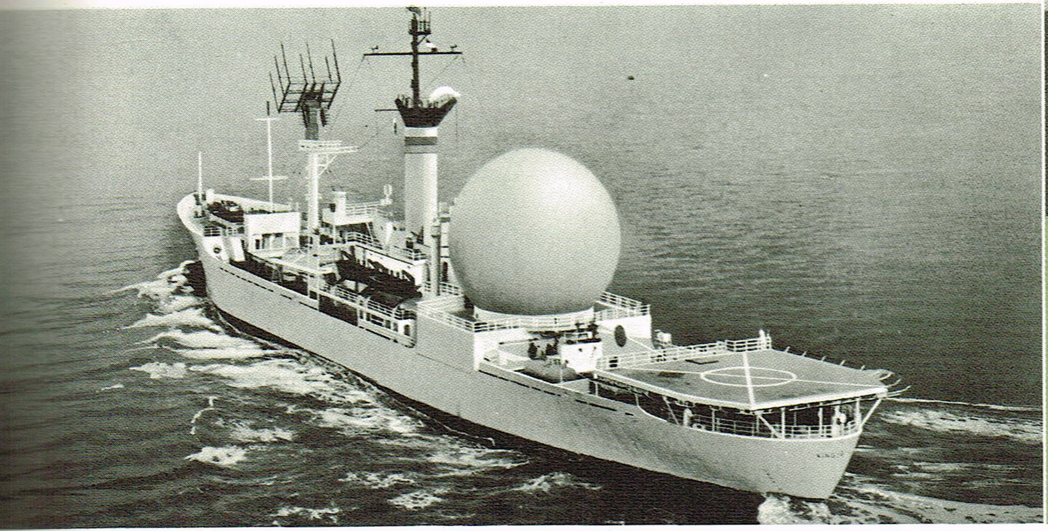
information

Autre mot magique. Que recouvre-t-il exactement? En vertu de quelle étonnante propriété, la « fée Électricité » — non contente d'être le véhicule le plus souple et le plus efficace de l'énergie et le carrefour de ses transformations — est-elle aussi le support le plus universel, sinon même l'unique support, de l'information?

On parle couramment de « traitement de l'information », comme au siècle dernier de « traitement des métaux ». Et les mots, d'apparence si barbare, « machines à traiter l'information » remplacent ceux, plus compréhensibles, de « machines électroniques à calculer », montrant à l'évidence — s'il en était encore besoin — combien étroitement sont liées électronique et information. L'imagination populaire s'enflamme, les auteurs de science-fiction nous proposent un monde de robots où les « cerveaux électroniques », rapides et sûrs, réduiront leurs anciens maîtres, véritables ilotes, en esclavage. Faut-il penser que l'homme est définitivement surclassé par les réalisations issues de son imagination? N'est-il déjà plus capable de suivre les progrès réalisés par des outils quelques milliards de fois plus puissants que ses... lointains ancêtres de 1943? Une vue plus raisonnable de



Relais hertzien et émetteur de télévision de Marseille.



Le « Kingsport » est, tout comme les stations terrestres du type Pleumeur-Bodou (v. p. 132), capable d'assurer les télécommunications *via* satellite.

Phot. U. S. I. S. et C. S. F., René Bouillot.

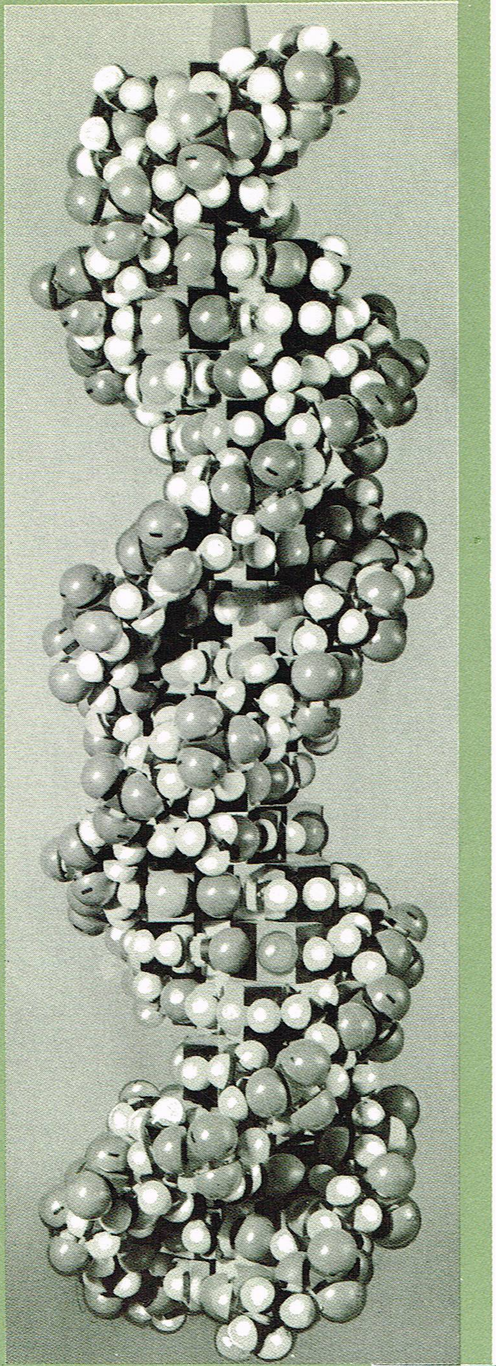
l'avenir est-elle possible, ramenant les robots à leur rôle de serviteurs? Extraordinairement puissants, sûrs et rapides, certes, mais incapables de toute initiative, réduits à l'exécution stricte des instructions reçues.

Il n'y avait pas de raison pour que la cybernétique, ou science de l'information (1947), ne tentât pas de s'appliquer aux composants mêmes de la vie; dès 1954, l'astronome (!) Gamov propose une première hypothèse pour déchiffrer le code même selon lequel les hélices d'acide désoxyribonucléique communiquent à d'autres macromolécules cellulaires l'information génétique qu'elles ont stockée, et qu'elles transmettent d'âge en âge. Ainsi s'ouvrent à l'homme des perspectives immenses, encore insoupçonnables, au cœur même des problèmes de la vie, de la biogenèse. Et probablement même de la cancérogenèse.

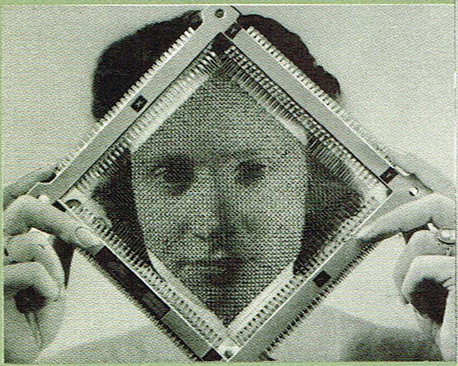
Comment toutes ces notions peuvent-elles se rattacher à l'aventure scientifique? Comment les théories abstraites et difficiles de la mécanique ondulatoire peuvent-elles expliquer le fonctionnement des instruments à semi-conduction? Et pourquoi ces derniers surclassent-ils les composants électroniques « classiques »? Comment l'optique, bien loin d'être la science morte que proposent encore certains livres d'enseignement, permet-elle l'exploration la plus fine de la matière inerte et de la cellule vivante? Quels sont exactement ces phénomènes de pompage optique et de cohérence de la lumière qui expliquent le fonctionnement des masers et des lasers? Ces derniers engendrent-ils seulement le « rayon de la mort » cher à une certaine forme de vulgarisation, ou bien n'ont-ils pas d'autres possibilités, beaucoup plus vastes? Comment peut-on mesurer l'information et montrer, comme le savent bien les statisticiens, que toute information, reçue ou transmise, est nécessairement limitée? Comment communiquer une si précieuse denrée, et comment la traiter le plus efficacement possible? Comment fonctionnent ces puissantes machines à faire des opérations logiques — et quelquefois même des opérations arithmétiques — à la cadence de plusieurs millions par seconde? Comment ont-elles rendu possible l'élaboration de cette discipline, jeune, complexe et si prenante, qu'est la « recherche opérationnelle » (ou science du gouvernement)? Signifie-t-elle que demain technocrates et peut-être robots nous dirigeront souverainement? Quelles perspectives sont ouvertes à l'industrie électronique? Est-il vrai que 50 p. 100 des appareils qui seront construits dans dix ans reposeront sur des découvertes, ou des brevets, encore inconnus? Est-il possible d'ignorer encore tant de choses après avoir amassé déjà un si riche butin de connaissances?

Nous essayerons, dans les pages à venir, de répondre à de telles questions. (L'étude de l'information génétique portée par les chaînes macromoléculaires d'acide désoxyribonucléique (A. D. N.) sera réservée à un chapitre ultérieur.)

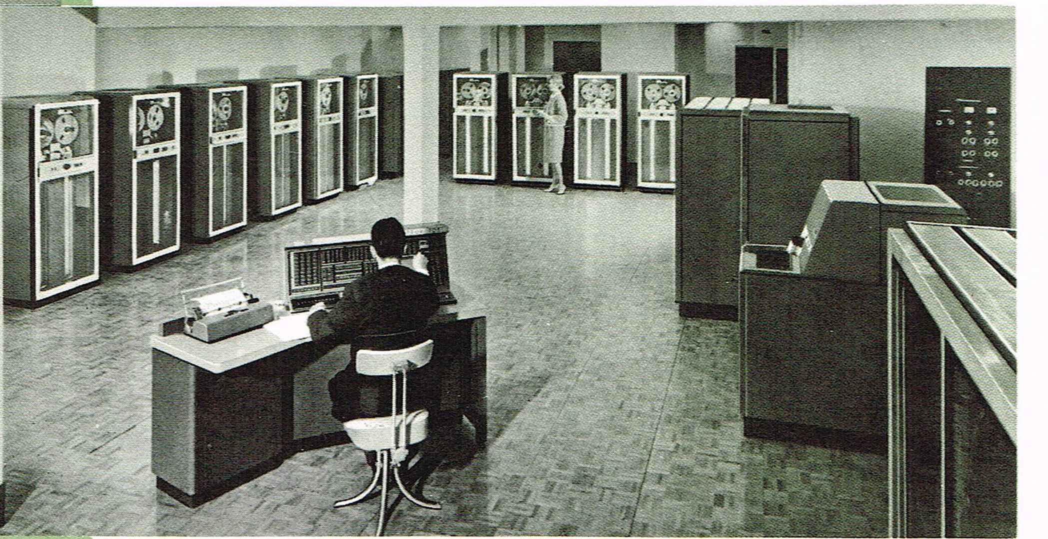
La lecture des chapitres suivants pourra être faite indépendamment de celle des autres, la plupart des notions théoriques étant présentées rapidement toutes les fois que le besoin s'en fait sentir. Le lecteur, et dans la seule mesure où cela est nécessaire, désireux d'acquérir une connaissance plus profonde des phénomènes de base de l'électronique est invité à se reporter aux chapitres du tome II, *Matière : ondes ou corpuscules* et *l'Architecture de la matière*, auxquels nous renverrons fréquemment.



Structure de l'A. D. N.



Grille à mémoire pour machine électrocomptable.



Ordinateur 705.

Phot. C. S. F., René Bouillot et IBM.

LES POSSIBILITÉS DE L'OPTIQUE MODERNE

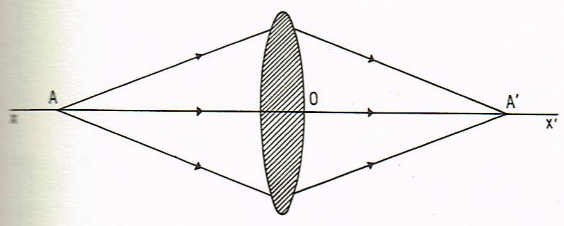
par

Maurice FRANÇON

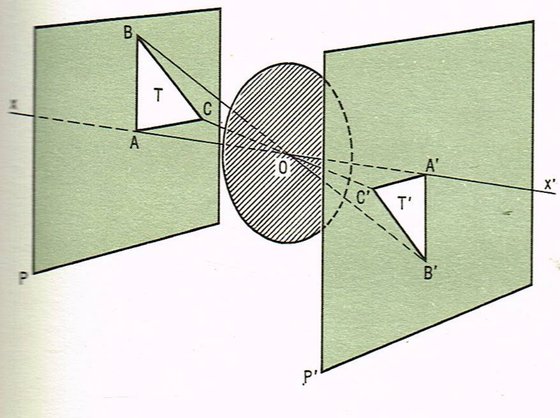
IL y a cinquante ans, les milieux scientifiques pensaient que l'optique était une science achevée et que son développement futur serait limité à des mises au point techniques. En fait, l'optique n'a pas cessé d'évoluer, et l'on assiste maintenant à un essor extraordinaire de cette discipline.

L'application à l'étude de la formation des images de méthodes mathématiques fondées sur la transformation de Fourier, le contraste de phase et le contraste interférentiel, les couches minces, les lasers sont autant de domaines dans lequels l'optique moderne a montré ce dont elle était capable. La conception que l'on se faisait des phénomènes a donc fondamentalement évolué, mais aussi les techniques, correspondant à d'innombrables et bien diverses applications; photographie, cinéma, télévision, biologie, médecine, chimie, métallurgie, cristallographie, astrophysique, recherches nucléaires et explorations spatiales ont bénéficié des progrès réalisés.

Il n'est pas possible d'aborder en détail un ensemble aussi vaste, et nous devons ici nous limiter à quelques exemples; puissent-ils montrer que, même à l'heure de l'atome, l'optique est toujours une science vivante et en plein développement.



A' est l'image géométrique de A et T' de T. Les contours de l'image sont parfaitement nets.

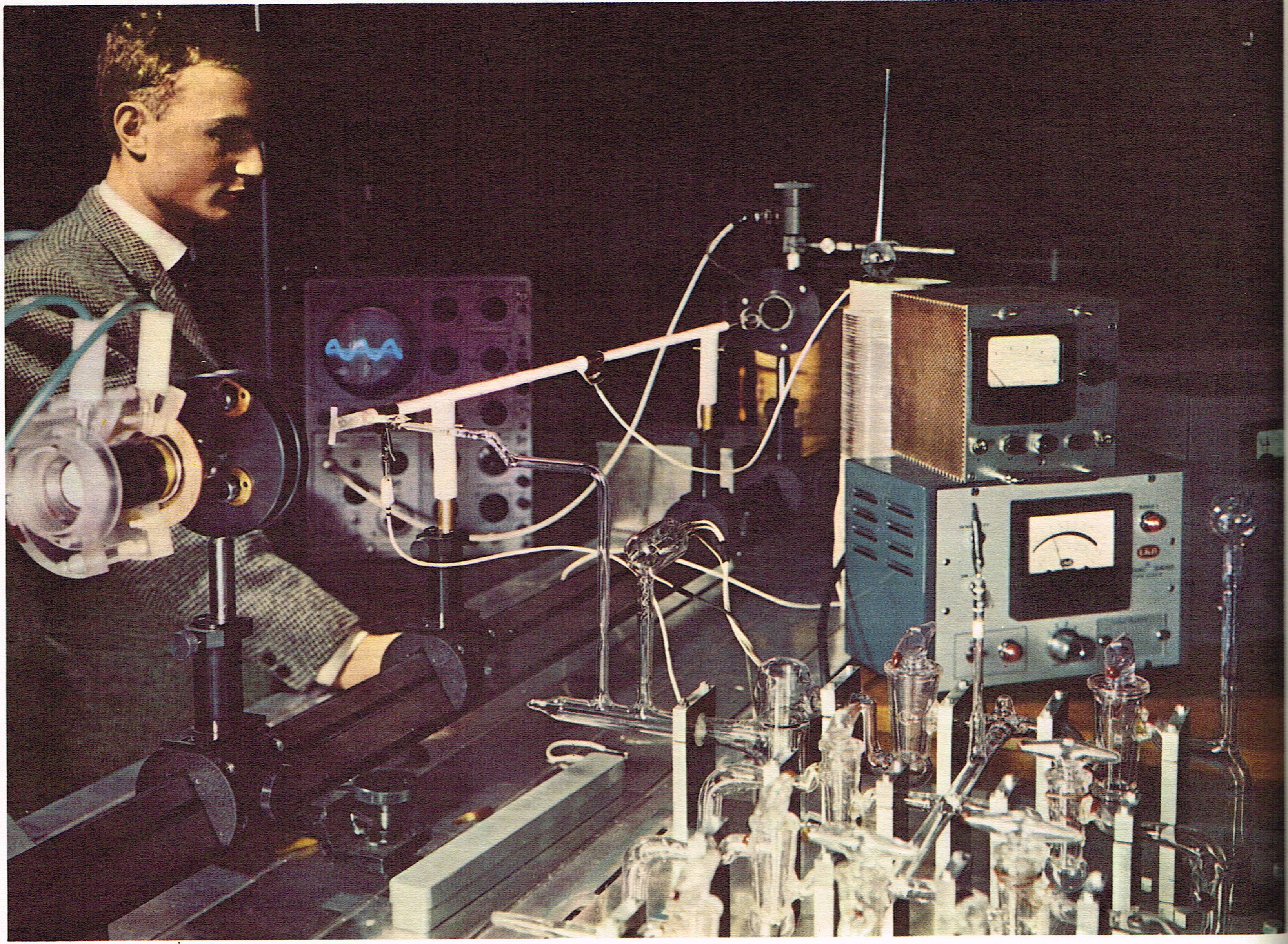


qu'est-ce qu'une image ?

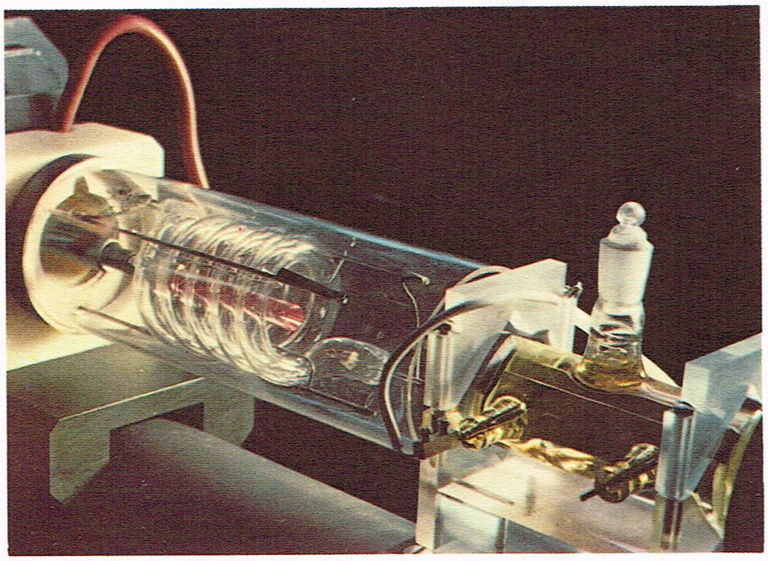
Lorsqu'on étudie du point de vue de l'optique géométrique l'image d'un objet donnée par un objectif, on considère d'abord le cas le plus simple, celui d'un objet « ponctuel ». On imagine des rayons lumineux partant d'un point A, le point objet; les rayons qui sont recueillis par l'objectif passent tous par le point A', image du point A. Remplaçons l'objet ponctuel par un objet étendu quelconque, par exemple un triangle lumineux ABC situé dans un plan P perpendiculaire à l'axe de l'objectif. A chaque point du triangle correspond un point image, et l'on obtient ainsi le triangle A'B'C', image de ABC, situé dans le plan P' perpendiculaire à l'axe optique. Si l'intérieur seul du triangle T est éclairé, il en est de même pour l'image : en dehors de T', il n'y a pas de lumière, et l'on passe brusquement de la région éclairée à la région obscure lorsqu'on franchit les côtés du triangle.

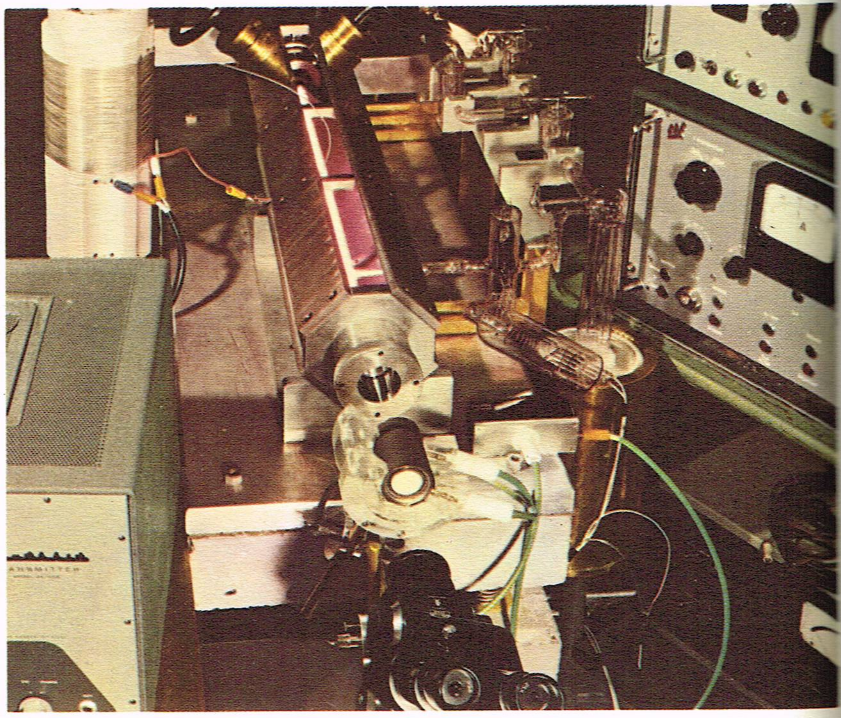
Réduisons les dimensions de l'objet : l'image diminue dans les mêmes proportions. Peut-il en être ainsi indéfiniment?

Certainement pas, et il est bien connu que les microscopes les plus puissants ne permettent pas de voir des objets aussi petits que les molécules. **Il y a une limite au-delà de laquelle l'optique géométrique ne permet plus d'expliquer la formation des images.** Cette limite fait intervenir des phénomènes liés à la structure de la lumière et appelés « phénomènes de diffraction », qui peuvent

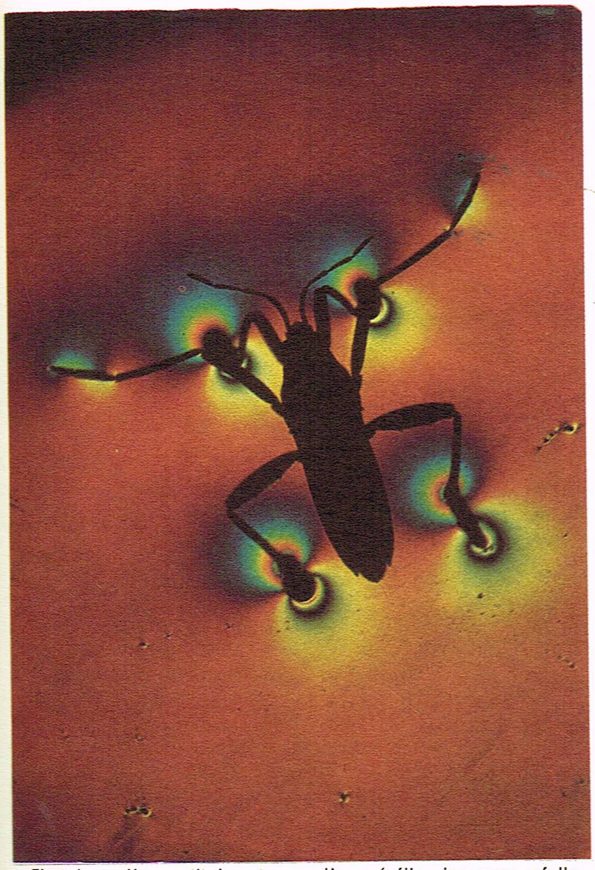


Lasers à gaz (*en haut et ci-contre*)
et à rubis (*ci-dessous*).

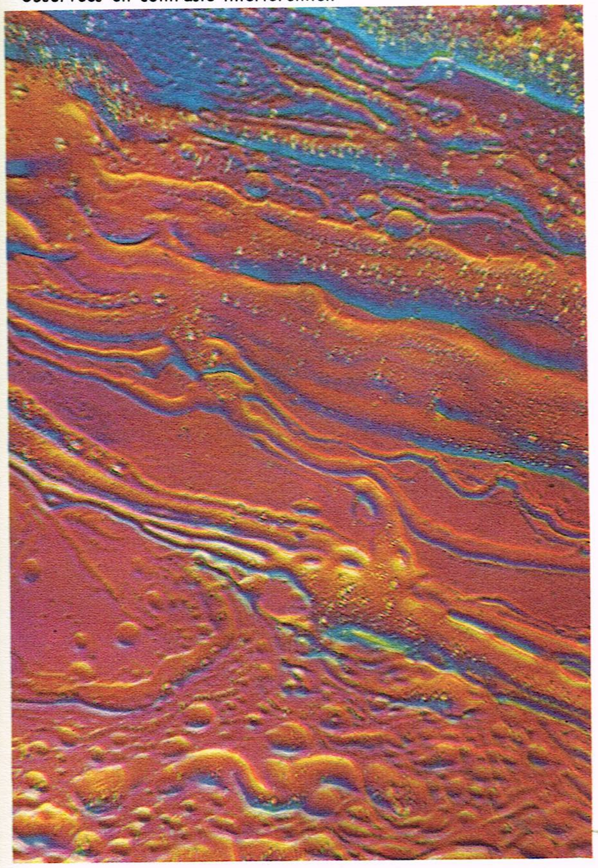




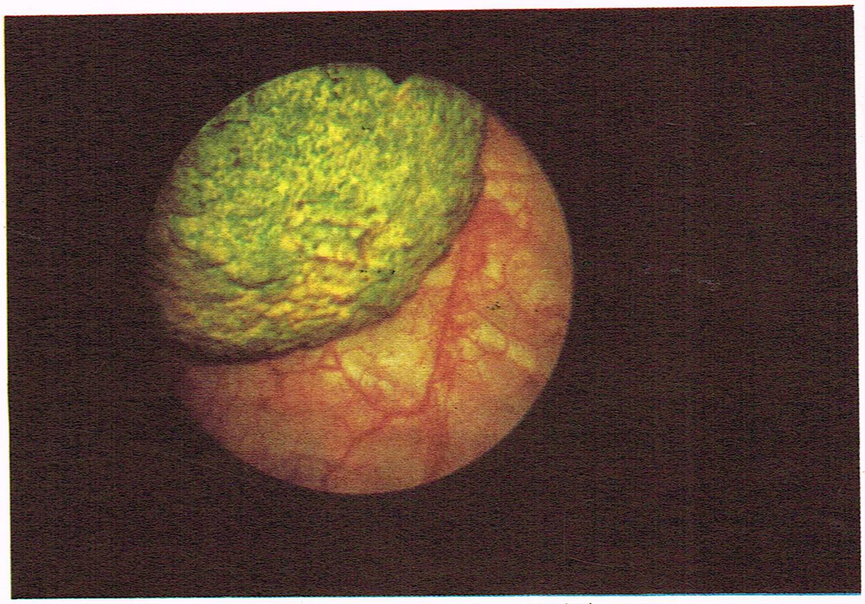
(*Phot. C. S. F. de R. Violet, J. C. Georgel et R. Bouillot.*)



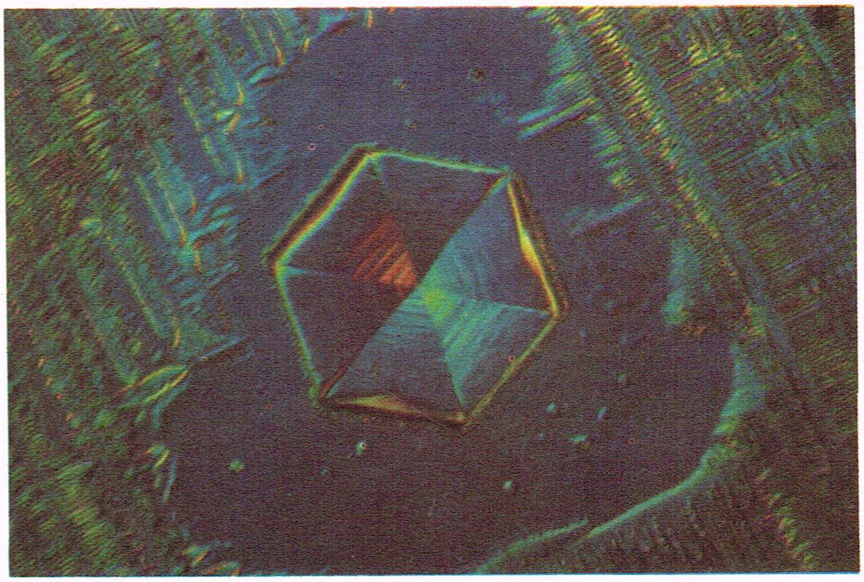
Flottaison d'un petit insecte sur l'eau (vélie, longueur réelle 7 mm). Les déformations de l'eau au contact de l'animal sont observées en contraste interférentiel.



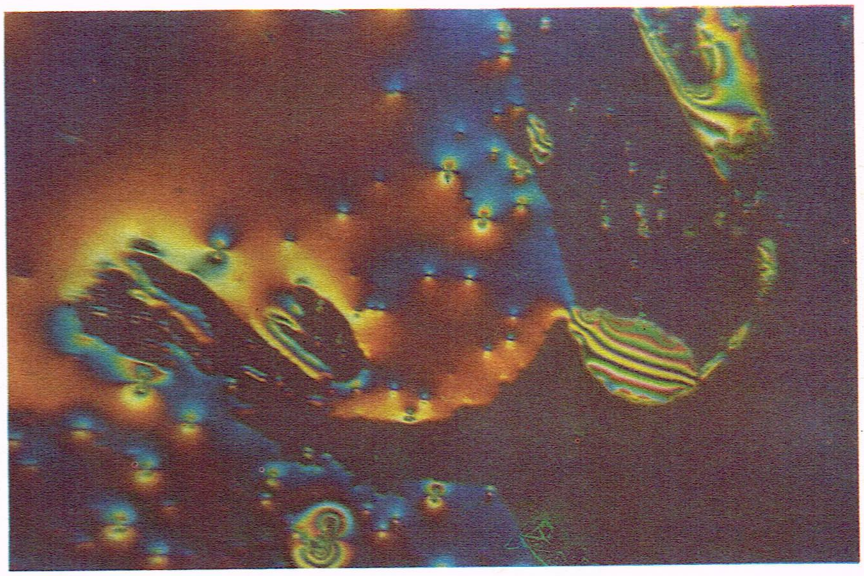
Variations de l'indice de réfraction d'un mélange de deux liquides transparents (contraste interférentiel).



Intérieur de vessie montrant un calcul (endoscopie du Dr Jaupitre).



Croissance d'un cristal d'alun d'ammonium (optiquement isotrope) en contraste interférentiel.



Variations d'épaisseur d'une goutte d'eau sur une lame de verre (contraste interférentiel).

(Phot. Institut d'Optique.)

s'expliquer par l'hypothèse de Fresnel suivant laquelle la lumière est une ondulation constituée (comme l'a montré Maxwell et comme nous le verrons longuement dans « la physique des ondes » [v. deuxième chapitre du tome II]) par des vibrations électromagnétiques, qui se propagent dans le vide, dans l'air et dans les milieux transparents.

l'image d'une source ponctuelle : une petite tache

L'hypothèse précédente implique que la lumière émise par A se propage sous forme d'ondes sphériques, de centre A (*fig.* 1). Pour un objectif parfait, les ondes transmises sont des sphères de centre A', image géométrique du point A. On peut admettre (principe de Huygens) que tous les points d'une surface d'onde, telle que Σ, jouent le rôle de petites sources lumineuses, réémettant des rayons : on dit qu'elles « diffractent » les vibrations.

Mais, attention, les vibrations émises par M_1 et M_2 ne sont pas indépendantes les unes des autres, et c'est là le point fondamental. On doit considérer que tous les points d'une surface d'onde sont dans le même état vibratoire.

Etudions ce qui se passe sur l'écran P' au point A', image géométrique du point A et centre de la sphère Σ (*fig.* 2). Les vibrations émises par M_1, M_2 et M_3 ont accompli le même chemin pour arriver en A' ($M_1A' = M_2A' = M_3A'$) : elles sont dans le même état en ce point et leurs effets s'ajoutent à chaque instant. Si l'angle α est petit, l'aspect des vibrations émises par les différents points de Σ est celui de la figure 3 : l'amplitude lumineuse en A' est la somme de toutes les amplitudes émises par tous les points. L'intensité lumineuse, proportionnelle au carré de l'amplitude totale, a donc sa valeur maximale en A'. On dit que les vibrations sont en « phase » en A'.

Considérons maintenant deux vibrations diffractées qui arrivent en un point B' quelconque (*fig.* 4[a]). Les vibrations, en phase au départ, n'arrivent plus nécessairement dans le même état vibratoire en B', car elles ont accompli un parcours différent : tout dépend de la différence $M_2B' - M_1B'$. Il peut arriver que la disposition des vibrations soit celle de la figure 4[b]. Les vibrations arrivent dans des états opposés en B' : leurs amplitudes sont égales et de signes contraires. L'amplitude en B' due aux points M_1 et M_2 est égale à la différence des amplitudes; si les deux points M_1 et M_2 étaient seuls à émettre des vibrations, il n'y aurait même pas de lumière en B'. On dit que ces deux vibrations sont en « opposition » en B'.

En fait, le point B' reçoit des vibrations diffractées de tous les points de la surface d'onde Σ; l'amplitude en B' sera donc obtenue en composant toutes les vibrations qui arrivent en ce point. Les vibrations émises par les points de l'onde Σ s'ajoutent, se retranchent et prennent toutes les dispositions possibles aux différents points du plan image P'. L'image du point lumineux A a une structure assez compliquée, même dans le cas simple où l'objectif est limité par un contour circulaire : on observe une tache centrale très lumineuse, entourée d'anneaux brillants diminuant d'ailleurs rapidement lorsqu'on s'écarte du centre A' (image géométrique de A). C'est la *tache de diffraction* (*fig.* 5).

Si λ est la longueur d'onde de la lumière utilisée et α l'ouverture de l'objectif, le rayon ρ de la tache centrale de diffraction nous sera donné par la formule simple

$$\rho = \frac{1{,}22\,\lambda}{\alpha}.$$

Pour un instrument d'ouverture $\alpha = 1/10$ utilisant la radiation verte du mercure ($\lambda = 0{,}546\ \mu$), le rayon de la tache de diffraction est $\rho = 6{,}7\ \mu$ ($1\ \mu = 10^{-3}$ mm).

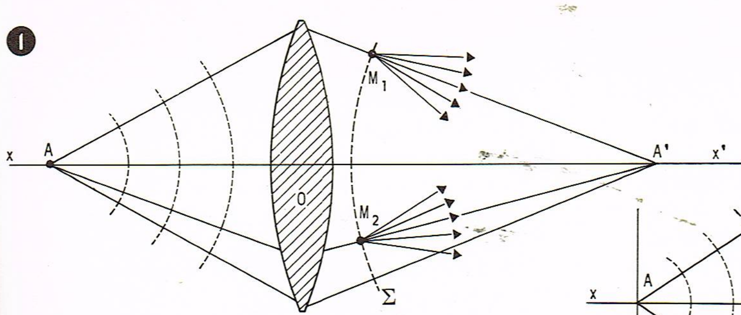
On voit que plus l'ouverture α de l'objectif est faible, plus la tache de diffraction s'agrandit. Il est donc avantageux d'employer des objectifs de grande ouverture, car ils donnent lieu à de petites taches de diffraction.

Si l'objectif est éclairé par deux sources lumineuses, il donne dans le plan image deux taches de diffraction; si les deux sources se rapprochent, il en est de même des deux taches de diffraction correspondantes. Il arrive un moment où A_1 et A_2 sont suffisamment proches pour que les deux taches A'_1 et A'_2 finissent par se confondre, et il est impossible de dire alors si l'on a affaire à deux sources A_1 et A_2 ou à une seule. Ainsi donc, **pour pouvoir séparer des points très rapprochés et percevoir des détails très fins dans une image, il faut réduire le plus possible la tache de diffraction et utiliser des objectifs de forte ouverture**. On admet que deux taches de diffraction sont séparées jusqu'au moment où leur distance est égale au rayon de leur tache centrale. C'est la notion traditionnelle du pouvoir séparateur d'un instrument d'optique.

la diffraction diminue la netteté des images

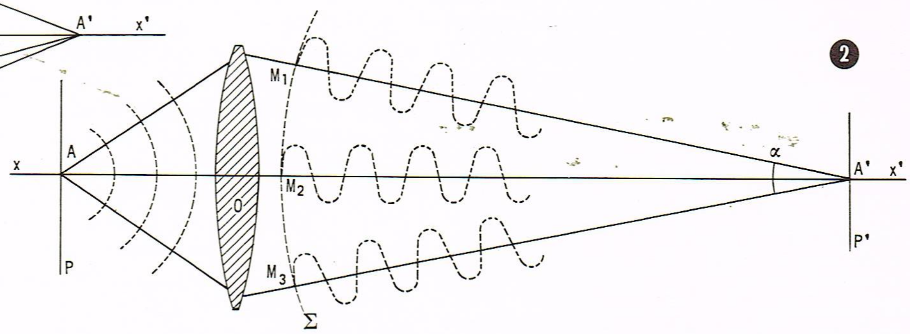
Nous avons parlé tout à l'heure de l'appareil parfait, cher à l'optique géométrique, donnant une image T' de l'objet T. Pour lui, les côtés A'B', B'C' et C'A' constituaient des limites parfaitement nettes entre les régions éclairées et les régions obscures du plan P'. **Par suite des phénomènes de diffraction, il ne peut pas y avoir de limite nette entre une région éclairée et une région obscure.** Puisque chaque point de l'image est une tache de diffraction, les points lumineux situés très près des bords du triangle A'B'C' doivent envoyer un peu de lumière en dehors de l' « image ». Les régions obscures et les régions éclairées sont séparées par des dégradés d'autant plus étalés que le diamètre de la tache de diffraction est plus grand.

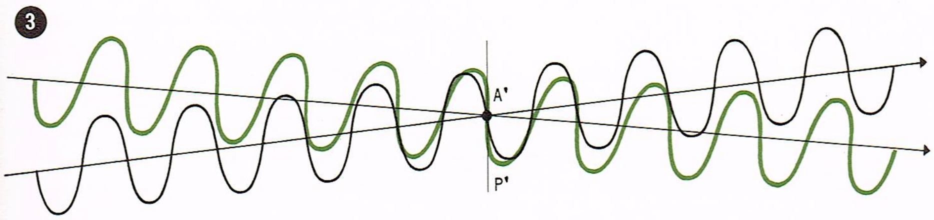
Phénomène peu important pour un objet de grandes dimensions; mais si les dimensions géométriques du triangle A'B'C' deviennent très petites, de l'ordre de grandeur de la tache de diffraction, il est facile de comprendre que l'image effective du triangle ABC peut être profondément modifiée par la diffraction. La tache de diffraction, image d'une source ponctuelle donnée par un instrument, joue donc un rôle considérable. En fait, les images d'objets étendus dérivent toutes de cette tache.



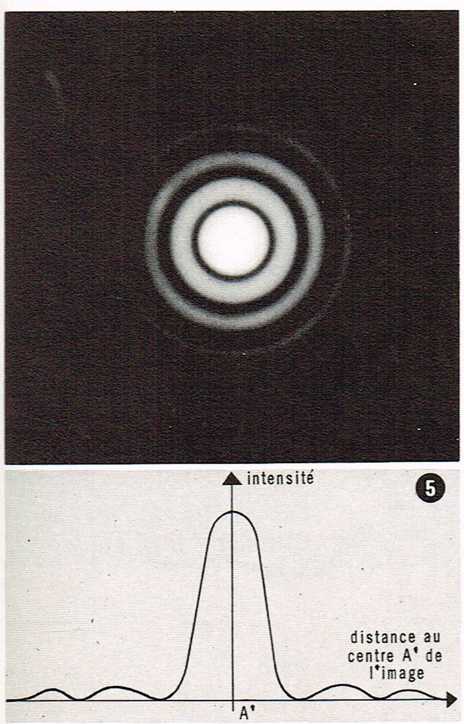
Le principe de Huygens : tous les points de la surface d'onde **Σ** émettent des vibrations en phase.

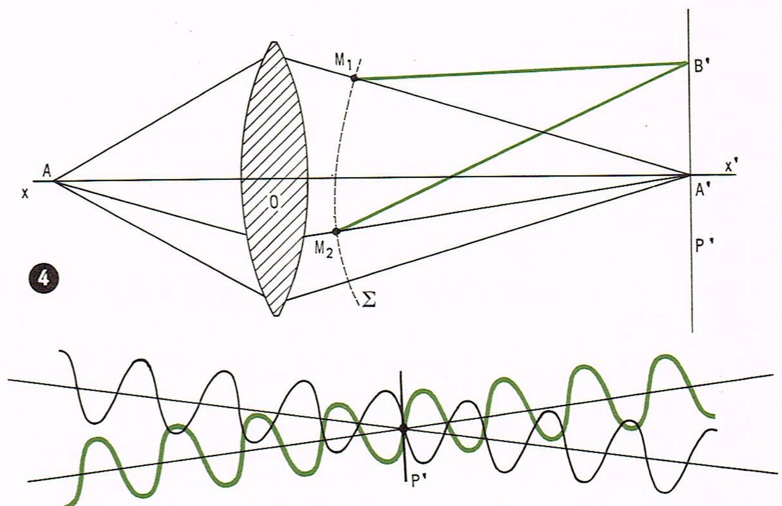
Seuls sont représentés ici les « rayons » M_1A' ... M_3A', qui transportent des vibrations en phase, ce que représente symboliquement la figure.





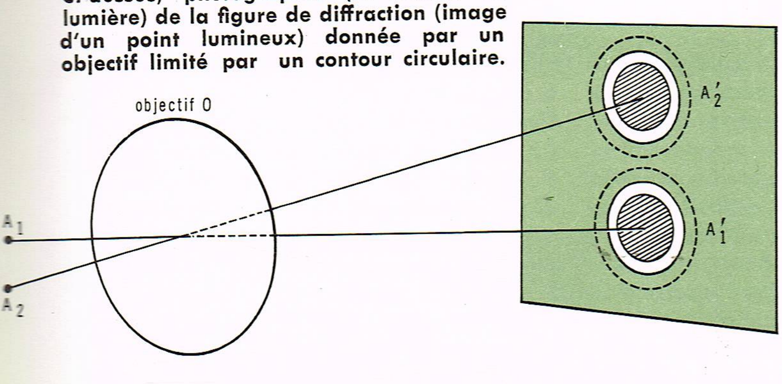
En A' les vibrations sont en phase, l'amplitude est la somme des amplitudes et l'intensité est maximale.

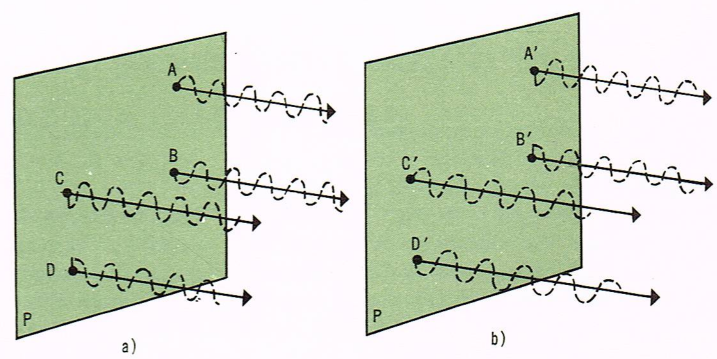




Si les chemins optiques sont différents, la différence de phase est quelconque; les vibrations issues de M_1 et de M_2 peuvent même être en opposition de phase et s'annuler : l'addition géométrique des vibrations lumineuses aboutit à l'obscurité.

Ci-dessus, photographie (et courbe de lumière) de la figure de diffraction (image d'un point lumineux) donnée par un objectif limité par un contour circulaire.





Un objet optiquement incohérent : les sources ABCD émettent indépendamment des vibrations sans relation entre elles, la disposition des sinusoïdes (pour simplifier nous n'avons ici représenté que les vibrations émises perpendiculairement à P) varie de façon quelconque lorsqu'on passe de A en B, en C...; il en est de même un instant plus tard lorsqu'on passe de A en A', de B en B', de C en C' ou de A' à B' ou C' ou D'.

objets incohérents et objets cohérents

On appelle objets optiques « incohérents » des objets dont les divers éléments émettent des vibrations parfaitement incohérentes, c'est-à-dire indépendantes les unes des autres.

Les objets de ce type sont extrêmement fréquents. Et tout d'abord les objets lumineux par eux-mêmes : étoiles, Soleil, sources terrestres variées, etc.; mais c'est aussi le cas des objets diffusants : planètes en astronomie, paysages, etc. De même les diverses radiations monochromatiques constituant le spectre d'une source sont incohérentes entre elles.

Prenons en revanche un objet éclairé par une source « ponctuelle » S placée au foyer d'un objectif C (*fig.* 1); les ondes sphériques de centre S sont transformées, après traversée de l'objectif C, en ondes planes (puisque l'image S′ est rejetée à l'infini). Interposons sur le faisceau des rayons parallèles un objet P, une photographie quelconque, par exemple. Nous supposons que les détails de la photographie sont constitués seulement par des variations d'absorption plus ou moins grandes et plus ou moins étendues. L'objet P ne présente aucune variation d'épaisseur. Lorsqu'une onde plane incidente telle que Σ atteint l'objet P tous les détails de la photographie « diffractent » un peu de la lumière incidente. La théorie et l'expérience montrent qu'en dehors d'un cône d'angle θ (d'autant plus grand que le détail considéré est plus petit) la lumière diffractée est pratiquement négligeable. Nous avons vu que tous les points d'une surface d'onde sont dans le même état vibratoire. Donc, les différents points de la photographie, éclairés en même temps par une même onde, diffractent des vibrations dans le même état vibratoire. On dit que les détails de la photographie P émettent des vibrations cohérentes et constituent un objet cohérent.

Il faut noter que la définition précédente reste valable même si les vibrations émises par A, B, C, etc., ne sont pas « en phase » (*fig.* 2). En effet, si l'objet P présente aussi de petites variations d'épaisseur, les vibrations émises par A, B, C, etc., ne sont plus dans le même état. La différence entre les états vibratoires de A, B, C, etc., reste cependant toujours la même au cours du temps, car les variations d'épaisseur de l'objet sont évidemment constantes.

La même figure peut représenter l'état optique d'un objet incohérent à un instant donné ou celui des vibrations émises par un objet cohérent. Dans ce dernier cas, la disposition des sinusoïdes les unes par rapport aux autres reste toujours la même; les vibrations peuvent se déplacer vers la droite, mais toutes de la même longueur. (Si A et B sont initialement en opposition, la disposition de leurs deux vibrations peut certes changer au cours du temps, mais dans un objet cohérent elles resteront en opposition.) Il en est de même pour tous les autres points de l'objet P. Un exemple type de l'éclairage cohérent se trouve réalisé dans le microscope lorsqu'on utilise une source d'éclairage très petite. Bien entendu, on peut situer entre l'incohérence et la cohérence toute une série d'éclairages intermédiaires appelés « éclairages partiellement cohérents », dont l'étude est très complexe. Nous savons que l'image d'une source ponctuelle est une tache de diffraction caractéristique de l'objectif utilisé, et nous venons de définir les deux types principaux d'objets étendus, c'est-à-dire non ponctuels, que l'on rencontre en optique. Nous pouvons donc étudier maintenant les images d'objets étendus, qui dérivent toutes de la tache de diffraction élémentaire.

un instrument d'optique est un véritable filtre

Les images A′, B′, C′ données par un objectif d'un objet incohérent sont des taches de diffraction de centres A′, B′, C′ (*fig.* 3). Les points A, B, C, émettent des vibrations incohérentes et l'on peut comparer les effets produits à ceux de très petites lampes électriques placées en A, B et C.

L'intensité lumineuse en un point quelconque est obtenue en faisant la somme des intensités produites séparément par les lampes. Dans le plan image P′, l'intensité en un point quelconque est la somme des intensités produites par les taches de diffraction.

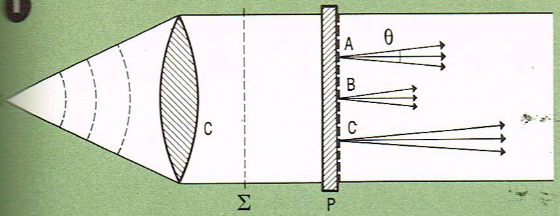
l'image d'une simple mire

Précisons l'objet en supposant qu'il s'agit d'une mire constituée d'un très grand nombre de traits noirs et blancs, parallèles et de même largeur : l'objet est périodique, de période *p*. D'après ce qui a été dit précédemment, la diffraction supprime les discontinuités entre régions éclairées et régions obscures de l'image.

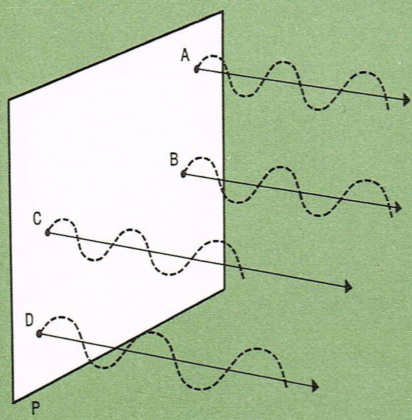
Tout point d'un trait blanc a pour image une tache de diffraction qui peut déborder sur un trait noir. Tout point d'un trait noir peut recevoir de la lumière, et l'on imagine aisément que la répartition de la lumière dans l'image de l'objet périodique doit être du type suivant.

La lumière envahissant l'image des traits noirs, ces derniers ont une intensité non nulle. Il y a maintenant un écart moins grand entre l'intensité des traits noirs et blancs de l'objet périodique. **Le contraste de l'image est inférieur à celui de l'objet.** Si la largeur des traits diminue, la tache de diffraction restant la même (on utilise toujours le même objectif), l'influence de la diffraction augmente de plus en plus et il arrive un moment où il n'y a pratiquement plus de contraste entre les traits : les traits sont trop serrés pour que l'objectif puisse en donner une image. **On a atteint la limite de séparation de l'objectif.** Il est très intéressant et très fécond d'étudier l'évolution de l'image de l'objet périodique en utilisant un théorème classique, dû à Fourier : toute courbe périodique quelconque peut être considérée comme résultant de la somme d'un nombre plus ou moins grand de sinusoïdes. L'objet étudié peut donc être représenté par la somme d'une infinité de sinusoïdes, d'amplitudes et de périodes déterminées.

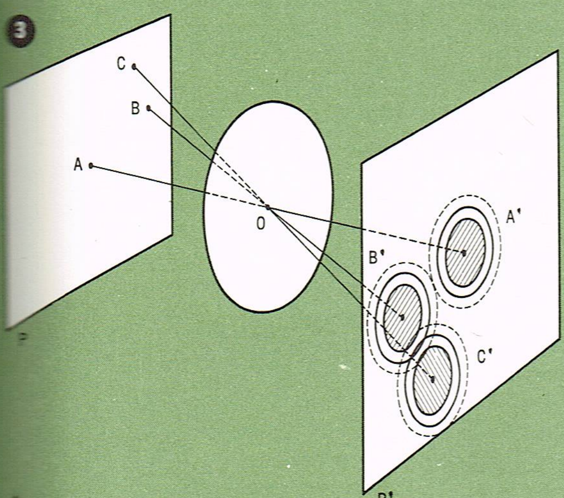
1

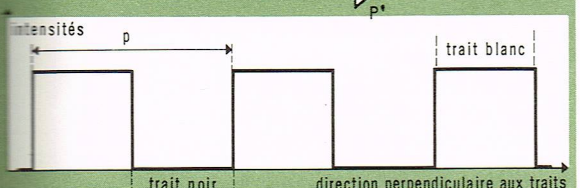


2

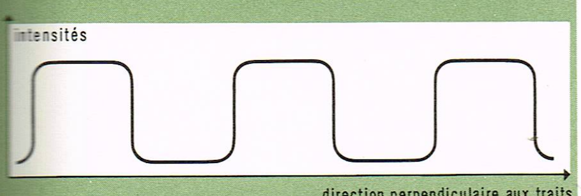


3





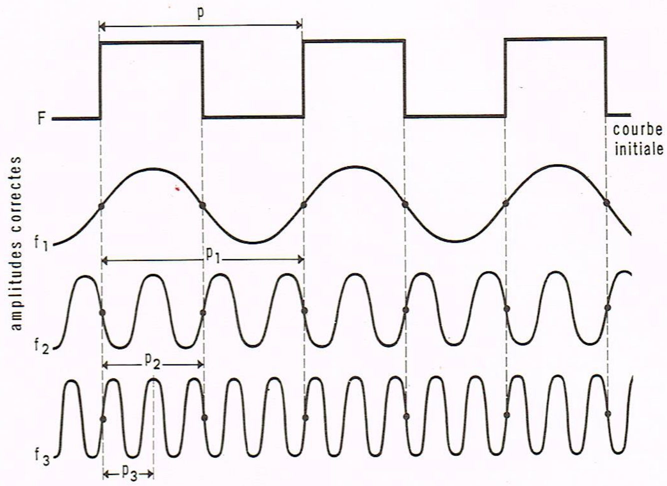
4



Une mire et la courbe d'intensité (périodique) définissant son image due à la diffraction (l'échelle des abscisses n'est pas la même).

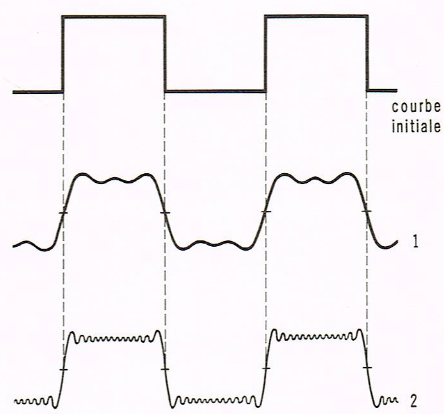
Un objet de période p (ou la composante de période p représentant la mire F) donne lieu à une meilleure image qu'un objet de plus courte période, puisque son contraste est moins affaibli par la diffraction; les composantes sinusoïdales de l'objet périodique ne sont pas transmises de la même façon par l'objectif. La composante « la plus large » est la mieux transmise (le contraste de son image est moins affaibli que celui des harmoniques suivants).

Ce point de vue conduit à transposer en optique le langage des électroniciens, et l'on considère une « fréquence spatiale » $1/p$ (inverse de la période p de l'objet). L'instrument d'optique va transmettre avec un « facteur » de transmission variable les composantes sinusoïdales des différentes fréquences. **Une fréquence basse (grande période, structure large) sera toujours mieux transmise qu'une haute fréquence (petite période, structure étroite).** L'instrument d'optique apparaît alors comme un filtre, et la formation de l'image est régie par une loi de filtrage des fréquences spatiales caractéristiques de l'instrument.



Le théorème de Fourier.

La fonction F est somme des fonctions sinusoïdales f_1, f_2, f_3,... (de périodes respectives p_1, p_2, p_3,...). *Ci-contre* (1), la somme des trois premiers « harmoniques » et celle (2) des dix premiers : l'écart entre cette seconde courbe et celle qui représente F est très petit et s'amenuiserait encore si l'on prenait un plus grand nombre de composantes.



la courbe de réponse caractérise un instrument

Ainsi donc, chaque sinusoïde composante est affaiblie lors de son passage à travers un instrument d'optique. Cela veut dire que le contraste de l'image d'une sinusoïde composante est toujours plus faible que celui de la sinusoïde objet. Caractérisons l'affaiblissement du contraste par un nombre T, qui caractérise la manière dont chaque composante sinusoïdale est transmise par l'instrument (pour $T = 1$ il n'y a aucun affaiblissement, et pour $T = 0$ l'affaiblissement est total : il n'y a plus d'image, le champ d'observation devenant uniforme). La courbe donnant T en fonction de p (v. illustration, p. 153) rencontre l'axe des x ($T = 0$). Toutes les fréquences spatiales supérieures à la fréquence $1/p_0$ ainsi définie sont coupées; $1/p_0$ est la fréquence limite, ou fréquence de coupure : **le récepteur ne permet pas de voir des structures dont la fréquence est supérieure à $1/p_0$, définie par $p_0 = \frac{\lambda}{\alpha} = 0{,}8\rho$** ($\alpha$ désigne toujours l'ouverture de

Phot. Larousse, C. S. F., IBM.

l'objectif). L'instrument est ainsi un filtre du type « passe-bas ». Cette fréquence limite suffit à caractériser complètement l'instrument, et c'est elle qui remplace la notion traditionnelle de « pouvoir séparateur ».

La définition que nous avons donnée de ce dernier (distance de deux points dont les taches de diffraction sont séparées d'une longueur égale au rayon du premier anneau noir) a manifestement un caractère conventionnel. Il est évident que l'on peut rapprocher les deux points tout en conservant une certaine séparation, car un observateur exercé reconnaîtra la présence de deux points au fait que la tache image n'est pas exactement circulaire. La notion de fréquence de coupure au contraire, associée à la connaissance complète de la loi de transmission en fonction de la fréquence, apporte toutes les informations utiles sur l'instrument d'optique.

On peut aussi calculer théoriquement la « courbe de réponse » d'un instrument d'optique, et la comparer avec la courbe obtenue expérimentalement; tout défaut de l'instrument se traduit par un écart entre les deux courbes : on peut ainsi « chiffrer » la qualité de l'instrument.

filtrage d'un objet tramé

On peut se servir de la courbe de réponse d'un instrument pour « filtrer » certains détails d'un objet (v. illustration, p. 153).

Prenons l'exemple simple d'une trame formée de petits points noirs. Dans les régions où l'objet est plus foncé, les points sont plus gros, et dans les régions claires, ils sont plus petits (les photographies sont en général reproduites en imprimerie par ce procédé). Supposons que l'objet à reproduire soit constitué par de larges bandes sombres (*fig.* 1); on distinguera alors deux types de fréquences : celle de la trame et celle des bandes sombres, la première évidemment beaucoup plus élevée que la seconde.

Formons une image de l'objet avec un objectif caractérisé par une courbe de réponse analogue à celle qui a été précédemment donnée (*fig.* 2^a^). L'objet est reproduit sur une feuille de papier blanc (*fig.* 3) et éclairé par diffusion (dans ces conditions l'éclairage est incohérent). Si la fréquence $1/p_1$ des bandes sombres et la fréquence $1/p_2$ de la trame sont inférieures à la fréquence de coupure $1/p_0$ (*fig.* 2^b^), l'objectif reproduit la trame et les bandes sombres, et l'image est semblable à l'objet (quoiqu'un peu moins contrastée). Si les fréquences précédentes sont telles que la fréquence $1/p_2$ de la trame est supérieure à la fréquence de coupure, la trame n'est pas transmise par l'objectif O (*fig.* 2^c^); en revanche, la fréquence $1/p_1$ des bandes sombres est inférieure à $1/p_0$, et l'objectif donne une image des bandes. On voit des bandes sombres plus ou moins floues se détachant sur un fond gris (*fig.* 4).

Comment peut-on agir sur la courbe de réponse pour obtenir un tel résultat? D'après ce que nous avons dit sur les phénomènes de diffraction, on voit que la courbe de réponse a une fréquence de coupure $1/p_0$ d'autant plus faible que la tache de diffraction est plus étalée. Or, pour augmenter le diamètre de la tache de diffraction, il suffit de réduire l'ouverture α de l'objectif. Dans ces conditions, en diaphragmant l'objectif, on peut obtenir une fréquence de coupure $1/p_0$ qui laisserait passer $1/p_1$ et couperait $1/p_2$.

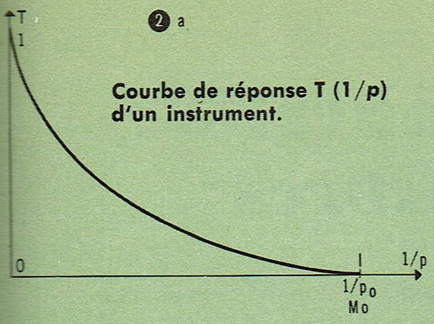
Mais, diaphragmer l'objectif, c'est réduire sa luminosité, ce qui est un inconvénient dans beaucoup de cas. On peut encore modifier la courbe de réponse en plaçant devant l'objectif O un écran transparent spécial, constitué par une série de très petites lames transparentes circulaires d'environ 0,6 μ d'épaisseur irrégulièrement disposées sur une lame de verre A servant de support. En plaçant A devant l'objectif O, on obtient le filtrage indiqué précédemment. Lorsqu'on reproduit par une trame une photographie déjà tramée, on obtient un effet très gênant : c'est le moiré classique. Le moiré a une fréquence plus grande que la plupart des détails de la photographie, et il est possible de l'éliminer avec le filtre précédent.

Etudions maintenant le filtrage des fréquences spatiales dans l'éclairage cohérent. Le problème est plus simple que dans le cas de l'éclairage incohérent. Considérons en effet une photographie éclairée par un faisceau parallèle (éclairage cohérent); les détails de la photographie sont constitués par des variations d'absorption qui diffractent la lumière. Nous avons représenté (*fig.* 5) la lumière diffractée par un détail A de la photographie. Elle est concentrée dans l'image A′ par l'objectif O. On peut constater que la lumière diffractée est très étalée dans le plan π passant par l'image S′ de S. Au contraire, le faisceau d'éclairage provenant de S est très étroit en S′. Cette remarque nous sera très utile non seulement pour étudier le mécanisme du filtrage dans l'éclairage cohérent, mais aussi plus tard, pour comprendre le principe même de la méthode du contraste de phase.

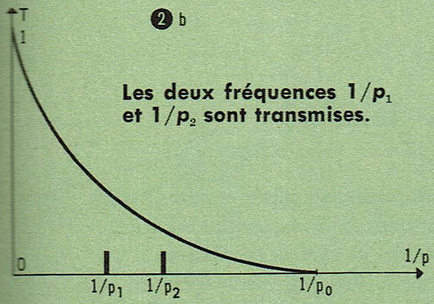
La lumière diffractée par A s'étale d'autant plus dans le plan π que ce détail est plus petit. Les détails de haute fréquence (petits détails) de la photographie P donnent une lumière diffractée très étalée en π. Au contraire, les détails de basse fréquence (détails plus larges) donnent une lumière diffractée qui s'écarte peu de S′. On imagine maintenant de quelle manière il faut agir pour filtrer les fréquences de l'objet P. Considérons de nouveau une trame, que nous éclairerons cette fois par de la lumière cohérente, arrivant en faisceau parallèle. La trame (haute fréquence) donne une lumière diffractée étalée en π; les bandes sombres (basse fréquence), une lumière diffractée resserrée autour de S′. Plaçons derrière l'objectif un écran opaque percé d'un trou de diamètre tel que toute la lumière diffractée par les bandes sombres le traverse. La lumière diffractée par la trame est arrêtée par l'écran opaque.

corriger le flou d'une photographie

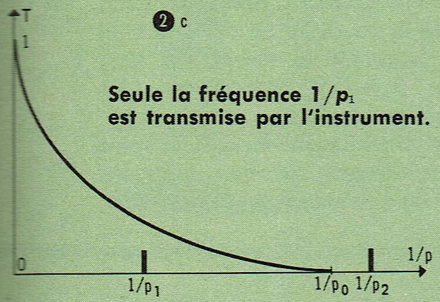
On peut utiliser un tel filtrage pour améliorer la qualité d'une photographie. Le bord des objets paraît d'autant plus net que la limite séparant deux régions d'intensités différentes est plus fine; si le bord des objets correspond à une discontinuité brutale, le bord est parfaitement net. Or, nous avons vu que les discontinuités sont d'autant mieux reproduites que l'instrument transmet plus de termes de haute fréquence. Donc, pour améliorer la netteté d'une photographie, il faut



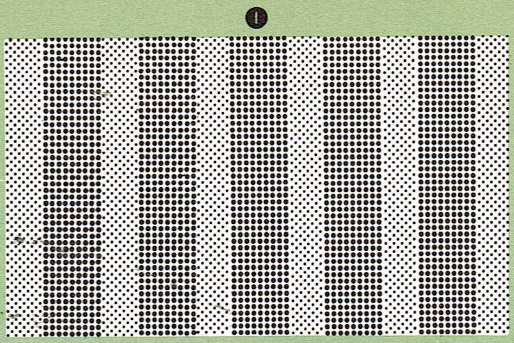
Courbe de réponse T (1/*p*) d'un instrument.



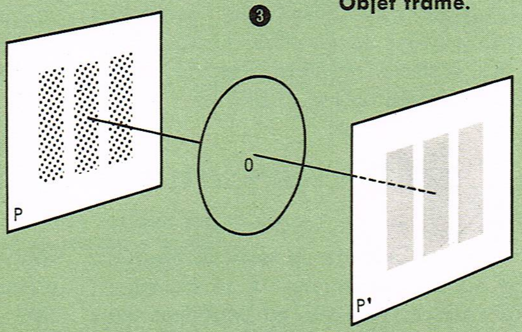
Les deux fréquences $1/p_1$ et $1/p_2$ sont transmises.

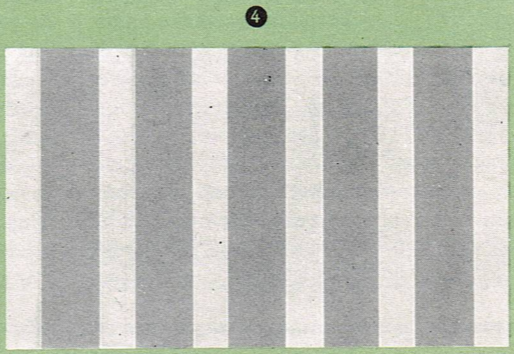


Seule la fréquence $1/p_1$ est transmise par l'instrument.

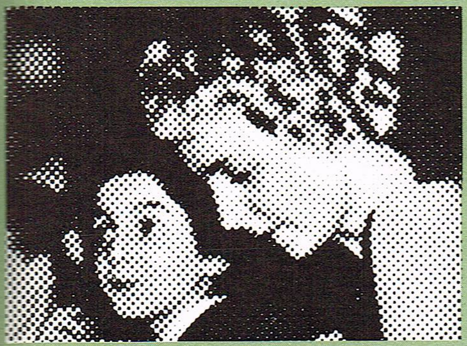


Objet tramé.



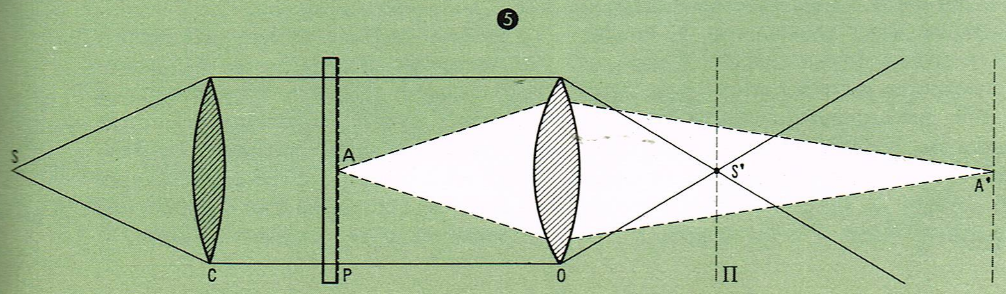


Objet tramé après filtrage.





Photographie tramée. *A droite,* la même image détramée; le filtre spécial a permis de faire disparaître la trame.

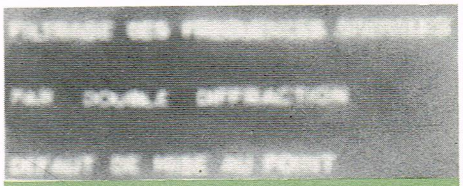


réduire la diffraction qui correspond aux basses fréquences (flou de la photographie) de façon à favoriser la diffraction par les termes de haute fréquence.

Examinons une photographie légèrement floue. On place sur l'image de la source une petite lame absorbante, assez étroite pour laisser passer les termes de haute fréquence donnant une diffraction étalée, mais néanmoins assez large pour absorber plus ou moins les termes de basse fréquence donnant une diffraction resserrée autour de S. La netteté de l'image en A′ est améliorée, ainsi que le montre la simple comparaison des trois photographies suivantes.



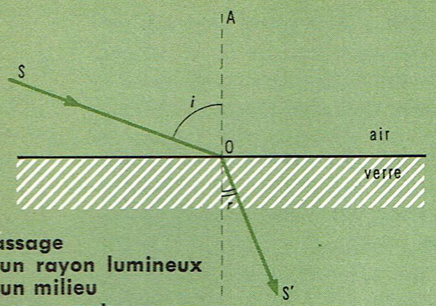
Photographie des lettres d'un texte.



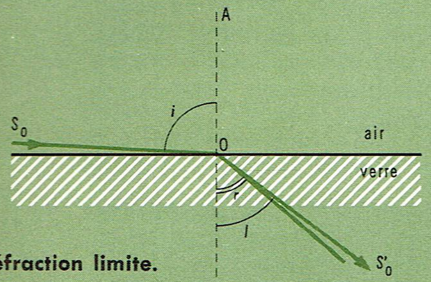
Un défaut de mise au point brouille les lettres.



Après filtrage, le flou est diminué et l'on peut reconnaître les lettres.



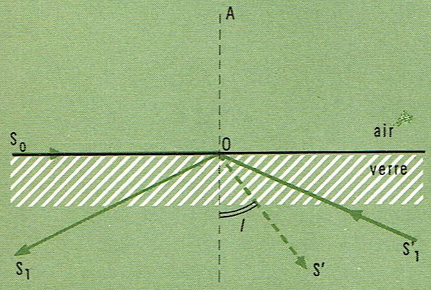
Passage d'un rayon lumineux d'un milieu dans un autre.



Réfraction limite.

Soit SO un rayon lumineux qui arrive en O sur une surface de verre plane (nous supposons que le milieu situé au-dessus de la surface de verre est l'air); il pénètre dans le verre, et le rayon OS' fait avec la perpendiculaire OA à la surface du verre un angle *r* plus petit que l'angle *i*. Renversons le sens de la lumière et supposons que le rayon se propage dans le sens S'OS : rien n'est changé et la disposition des rayons reste la même que sur la figure. Considérons un rayon incident SO de plus en plus incliné; l'angle *r* augmente de plus en plus jusqu'à la valeur *l* correspondant à un rayon S_0O rasant. En renversant le sens de la lumière, on peut dire qu'au rayon S'_0O correspond un rayon OS_0 confondu avec la surface.

Considérons maintenant un rayon S'_1O faisant avec OA un angle supérieur à l'angle limite *l*. Il ne peut y avoir de rayon émergent, et l'expérience montre que le rayon S'_1O se réfléchit en O comme sur un miroir parfaitement réfléchissant. On dit que le rayon S'_1O subit la réflexion totale. On retrouve dans le rayon réfléchi OS_1 toute l'énergie du rayon incident S'_1O. Pour que ce phénomène se produise, il faut que la réflexion se fasse d'un milieu plus dense, par exemple le verre, sur un milieu moins dense, par exemple l'air. Il faut aussi que le rayon incident S'_1O fasse avec OA un angle plus grand que l'angle *l*, appelé « angle limite de réflexion totale ».



une optique nouvelle : l'optique à fibres

Les instruments d'optique classiques sont formés de lentilles ou de miroirs : les rayons issus d'un point de l'objet convergent en un point de l'image grâce à des surfaces réfléchissantes ou réfractantes placées entre l'objet et l'image. Il est possible de former l'image d'un objet par un mécanisme différent fondé sur le phénomène de la réflexion totale, qui est à la base des instruments d'optique à « fibres ».

les ondes prisonnières des fibres

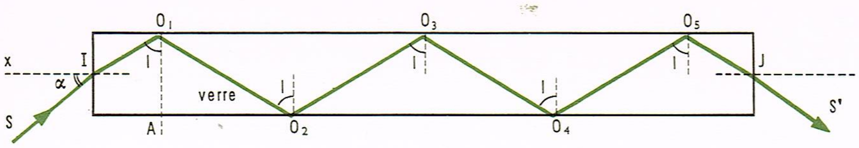
Prenons un cylindre de verre de longueur très grande par rapport à la section. Il existe des rayons qui, en pénétrant dans le cylindre, subissent en O_1 la réflexion totale. Toutes les réflexions suivantes, en O_2, O_3, ..., se font sous le même angle, et le rayon chemine dans le cylindre de verre par réflexions totales successives sans perdre d'énergie. Finalement, il sort suivant JS'. Tous les rayons incidents formant avec l'axe un angle plus petit que α (correspondant à une réflexion limite en O_1) subiront des réflexions totales dans le cylindre de verre. Donc, les rayons situés dans un cône d'angle 2 α émergeront tous du cylindre après de multiples réflexions totales. Si l'on néglige l'absorption du verre et les imperfections du cylindre, tout le flux lumineux incident se retrouve à la sortie du cylindre. **Le cylindre de verre se comporte comme un « conducteur » de lumière**. Le phénomène n'est pas changé si l'on courbe le cylindre de verre, pourvu que la courbure ne soit pas trop forte. On reconnaît le mécanisme bien connu des fontaines lumineuses : les veines liquides conduisent le flux lumineux de la même façon; lorsque la veine liquide se disloque, le flux lumineux s'éparpille en produisant de beaux effets.

un faisceau de fibres peut transporter une image

Etudions le mécanisme de la formation des images à l'aide des mêmes cylindres que les précédents, mais de très petit diamètre, par exemple de l'ordre de 0,05 millimètre (50 μ), véritables fibres qui peuvent être faites en verre ou en matière plastique.

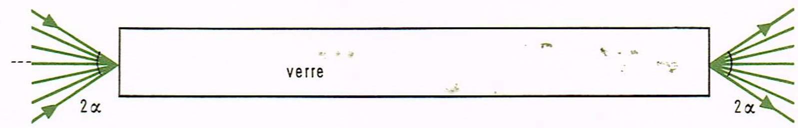
Groupons un très grand nombre de fibres identiques et parallèles en faisceau, et appliquons un objet plan contre l'une de ces faces bien planes. L'objet est constitué, par exemple, par une plaque photographique, et nous supposerons pour commencer qu'elle ne comporte qu'un seul détail : une petite tache noire en A, dont le diamètre est de l'ordre de celui des fibres. La plaque P étant éclairée, toutes les fibres transportent la lumière qui se trouve devant chacune d'elles par le mécanisme précédemment décrit. Seule la fibre qui se trouve devant A ne transporte pas de lumière, puisqu'elle n'est pas éclairée. Plaçons sur la face terminale du faisceau de fibres une plaque photographique vierge P'. La plaque P' est impressionnée à travers toutes les fibres, sauf à travers celle qui, partant de A, arrive en A'. Remplaçons P par une plaque comportant un détail plus grand, par exemple un carré noir dont le côté serait égal à cinq ou six fois le diamètre des fibres. Toutes les fibres qui se trouvent placées contre le carré ne sont pas éclairées, mais toutes les autres transportent de la lumière, qui impressionne la plaque. L'image obtenue sur la face terminale du faisceau sera d'autant plus fine que le diamètre des fibres sera plus petit par rapport aux dimensions des détails. Le faisceau de fibres « conduit » l'image du carré.

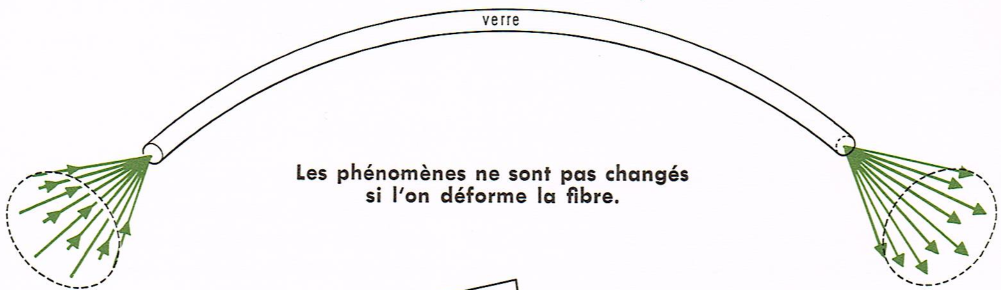
Le faisceau de fibres se comporte comme un instrument d'optique qui transporterait la lumière d'un objet P quelconque en une image P' dont les dimensions seraient les mêmes que celles de l'objet. Un faisceau souple comprenant un grand nombre de fibres peut être muni d'un objectif à une extrémité et d'un oculaire à l'autre. Il



Propagation d'un rayon dans un cylindre de verre.

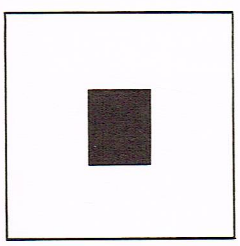
Le cône de lumière à l'entrée
se retrouve à la sortie.

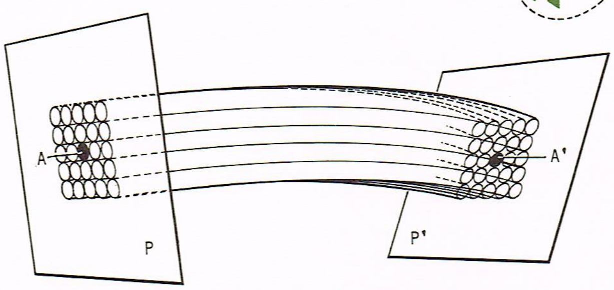


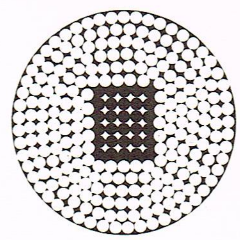


Les phénomènes ne sont pas changés
si l'on déforme la fibre.

L'image est d'autant
meilleure que le
diamètre des fibres
est plus petit par
rapport aux dimensions
du carré.

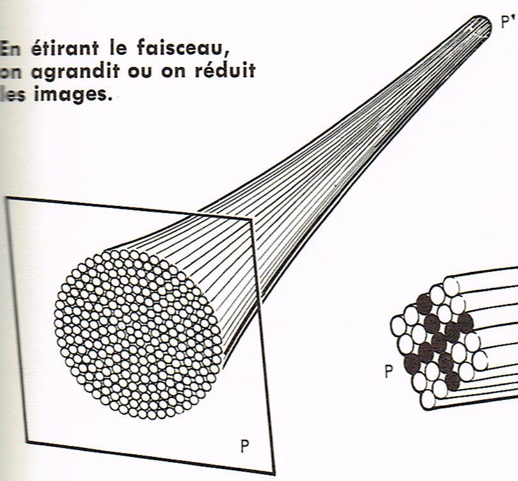






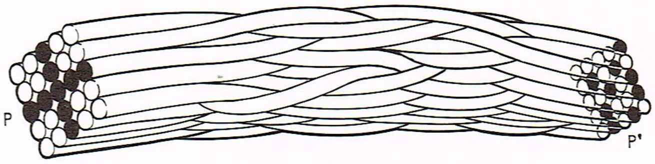
Objet noir
de forme carrée.

Un faisceau formé de fibres identiques
se comporte comme un instrument d'optique
donnant une image égale à l'objet.

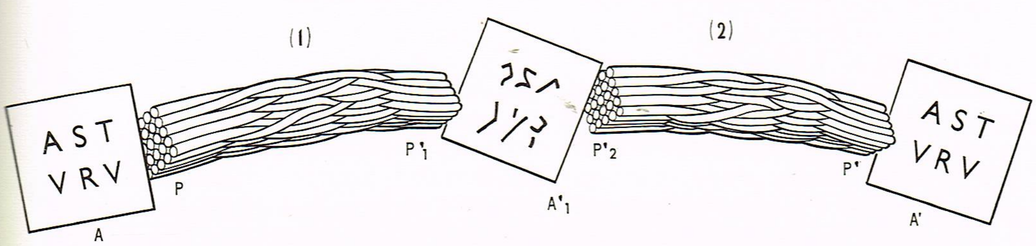


En étirant le faisceau,
on agrandit ou on réduit
les images.

Si les fibres sont mélangées,
on peut déformer les images.



L'image est déformée... mais on peut la reformer.



permet de voir dans des directions très différentes et constitue un véritable « œil sur pédoncule ».

On utilise des faisceaux souples de fibres en médecine, pour éclairer ou observer les cavités humaines **(endoscopes).** Au lieu de donner au faisceau des faces terminales planes, on peut donner à la face d'entrée P une forme incurvée qui épouserait au mieux le champ image d'un objectif. En laissant l'autre face plane, on a là un procédé simple de correction de la courbure du champ.

Il est possible d'obtenir des faisceaux de fibres qui transportent des images en les grandissant. Etirons l'extrémité d'un faisceau de fibres par chauffage. Le diamètre de chaque fibre est ainsi diminué dans un rapport déterminé, et, par conséquent, les dimensions de l'image aussi. Si l'objet est en P, l'image en P′ est réduite, mais on peut mettre l'objet en P′, et l'on obtiendra une image agrandie en P. On peut également modifier la symétrie des faisceaux, et obtenir des applications intéressantes en spectrographie astronomique.

pour transmettre des messages secrets

V. figure en bas de la p. 155.

Une déformation complète d'un faisceau de fibres peut servir à coder des messages. Considérons un faisceau dans lequel toutes les fibres sont complètement mélangées, mais dont la disposition est rigoureusement la même sur les faces terminales. En plaçant en P un objet, on trouve en P′ une image identique à l'objet, car, si les fibres effectuent des trajets compliqués entre P et P′, elles retrouvent en P′ une disposition identique à celle qu'elles avaient en P. Coupons le faisceau en deux. Deux personnes qui désirent correspondre possèdent l'une le faisceau 1 et l'autre le faisceau 2. Toutes les faces P, P'_1, P'_2, P′ sont bien planes et, par construction, la disposition des fibres est donc la même en P et en P′. Malgré le mélange des fibres, leur disposition est la même en P'_1 et en P'_2, puisque ces deux plans proviennent de la même section. Plaçons contre P un objet A, le message à transmettre. En P'_1, nous trouvons une image A'_1 complètement déformée par suite du mélange des fibres. C'est l'image qui est envoyée à l'autre correspondant. Celui-ci place le message incompréhensible A'_1 contre la face P'_2 de son faisceau. En P′ celui-ci restitue le message initial. Si l'on imagine des faisceaux parallélépipédiques de 5 cm de côté avec des fibres de 0,05 mm, donc constitués d'un million de fibres, on pourrait transmettre des messages très riches, et dont le déchiffrage serait pratiquement impossible par celui qui ne disposerait du code.

l'exploration interne du corps humain

L'emploi d'un grand nombre de fibres est nécessaire pour transmettre des images, mais l'on peut se contenter de quelques fibres, voire d'une seule, dans un autre dessein : celui d'éclairer un objet inaccessible. Nous revenons au principe même du fonctionnement d'un cylindre de verre. En négligeant les pertes par absorption, pratiquement tout le flux lumineux qui entre dans le cylindre se retrouve à la sortie. Cette propriété est particulièrement utile dans l'endoscopie médicale, lorsqu'on veut observer des cavités d'accès difficile comme, par exemple, les bronches. Pour pouvoir photographier ou cinématographier les bronches, il faut beaucoup de lumière. Les fibres apportent une solution élégante à ce problème; une lampe extérieure au patient fournit le flux lumineux nécessaire, qui est transmis par elles dans les cavités à examiner. Dans un « bronchoscope » de ce type, le cylindre a 2 mm de diamètre environ et 30 cm de longueur. Des résultats remarquables ont aussi été obtenus dans l'observation du cerveau.

observer un objet normalement invisible

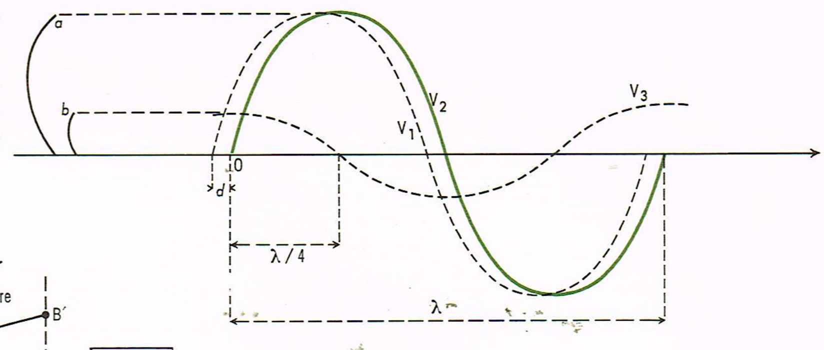
Pour être perçu par l'œil humain, un objet ne doit avoir ni la même couleur ni la même intensité lumineuse que le champ dans lequel il se trouve. Lorsque les différences avec le milieu ambiant sont trop faibles, l'objet est invisible par manque de contraste. On ne peut distinguer les détails d'un paysage que grâce aux différences de couleurs ou d'intensité qu'ils présentent. Par temps de brume. les mêmes détails s'estompent et finissent par disparaître si leur contraste devient inférieur à une certaine valeur critique.

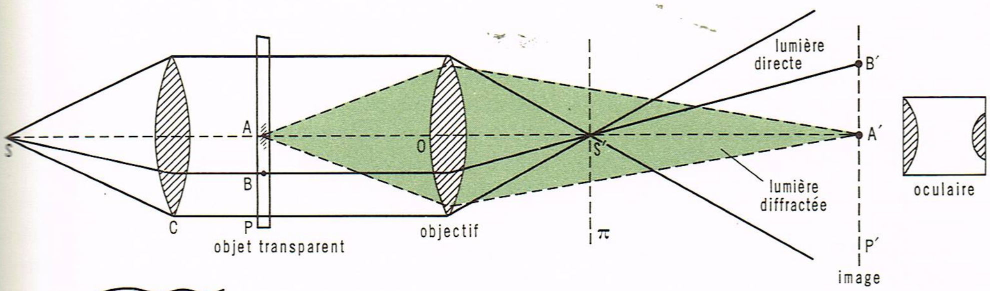
En biologie, par exemple, la plupart des préparations microscopiques sont des préparations transparentes. L'objet et les fins détails qu'il présente ne sont caractérisés que par de faibles variations d'indice de réfraction ou d'épaisseur; aucune coloration n'y apparaît. Tel est le cas d'une petite bactérie transparente, dont l'indice de réfraction est peu différent du liquide, également transparent, dans lequel elle baigne. Tout le champ d'observation est uniformément éclairé et aucun détail n'est visible. **C'est la raison pour laquelle les biologistes provoquent artificiellement des contrastes par l'emploi des colorants, qui malheureusement tuent les objets vivants.**
Cette classe des objets transparents joue un rôle important non seulement en biologie, mais aussi en médecine, en physique, en chimie, etc. Ces objets sont invisibles non pas parce qu'ils sont trop petits, mais parce qu'ils manquent de contraste. La méthode du contraste de phase imaginée par le physicien hollandais Zernike rend visibles de tels objets. Ultérieurement, le contraste interférentiel est venu apporter encore d'autres éléments d'information dans la vision des objets transparents. Contraste de phase et contraste interférentiel sont parmi les perfectionnements les plus remarquables mis au point depuis la dernière guerre dans l'optique.

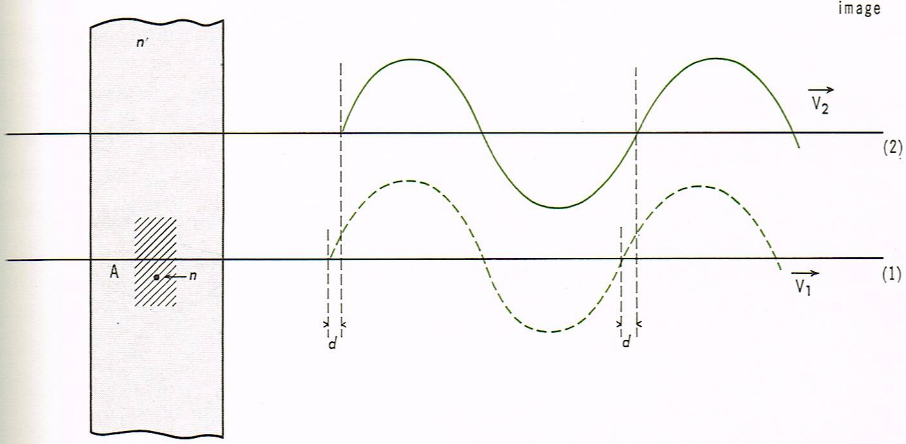
le contraste de phase permet de voir des objets transparents

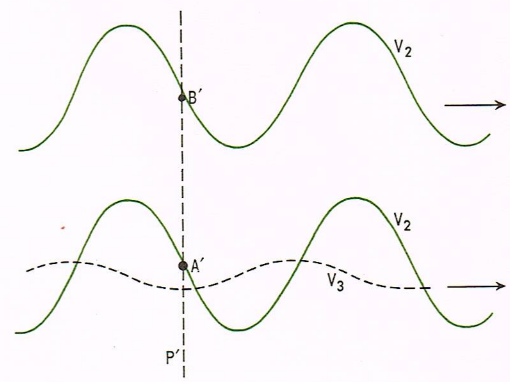
Le schéma ci-contre, bien connu, permet d'expliquer le principe de la méthode du contraste de phase. Nous allons choisir un objet transparent très simple, par exemple une lame de verre à faces parallèles dont l'indice de réfraction est constant dans toute la lame, sauf dans une petite région A. Aucune variation d'intensité lumineuse ne permet de distinguer cette région du reste de la lame, et l'objet apparaît uniformément transparent; la région A, ne présentant aucun contraste avec le reste de la lame, ne peut être visible. Sous l'action du faisceau de rayons parallèles qui l'éclaire, la région A diffracte un peu de la lumière incidente. La plus grande partie de la lumière traverse P comme si l'indice de réfraction était parfaitement constant dans toute la lame; l'objectif reçoit un faisceau de lumière parallèle; les rayons de lumière directe passent par le foyer de l'objectif, c'est-à-dire par l'image de la source, et s'étalent

$\vec{V}_1$ étant très peu différent de $\vec{V}_2$, et de même amplitude, on peut écrire $\vec{V}_1 = \vec{V}_2 + \vec{V}_3$.
$\vec{V}_2$ représente la lumière directe du champ, $\vec{V}_3$ la lumière diffractée, son amplitude est nulle pour $d = 0$, et d'autant plus petite que le retard d est faible. On constate et on vérifie aisément par le calcul que $\vec{V}_3$ est décalé par rapport à $\vec{V}_1$ ou à $\vec{V}_2$ d'un quart de longueur d'onde.









dans le plan image P′. On peut constater que le faisceau de lumière directe (provenant de S) se sépare nettement en S′ du faisceau de la lumière diffractée par A. Dans le plan passant par S′, la lumière directe est très resserrée en S′, et la lumière diffractée très étalée.

Etudions la région A où se trouve la variation d'indice de réfraction. Si son indice de réfraction est un peu plus grand que l'indice du reste de la lame, les vibrations (rayon 1) qui traversent A sont légèrement en retard sur les vibrations (rayon 2) qui traversent la lame en dehors de A. Puisque la transparence de la lame P est la même partout, les deux vibrations $\vec{V}_1$ et $\vec{V}_2$ sont de même amplitude; elles sont simplement décalées l'une par rapport à l'autre d'une petite longueur d. Représentons les vibrations $\vec{V}_1$ et $\vec{V}_2$ sur la même figure. Nous avons dessiné une troisième vibration $\vec{V}_3$, dont l'amplitude b est beaucoup plus petite que l'amplitude a des vibrations qui traversent P. Que représente cette nouvelle vibration $\vec{V}_3$? Un simple examen de la figure montre que la somme des vibrations $\vec{V}_2$ et $\vec{V}_3$ redonne la vibration $\vec{V}_1$. Plus le décalage des vibrations $\vec{V}_1$ et $\vec{V}_2$ est faible, plus l'amplitude b de la vibration $\vec{V}_3$ est petite, comme il est facile de le voir. Or, le décalage d est d'autant plus faible que la différence des indices est plus petite.

La vibration $\vec{V}_3$ disparaît si la région a le même indice que la lame : elle représente la lumière diffractée par la région A.

La vibration $\vec{V}_1$ qui traverse A peut donc être remplacée par la somme des vibrations V_3 (vibration diffractée) et V_2 (vibration directe). On constate que la vibration $\vec{V}_3$ est fortement décalée par rapport aux autres. Si le décalage d reste toujours petit, ce que nous avons supposé, le décalage de la vibration $\vec{V}_3$ par rapport aux vibrations $\vec{V}_1$ et $\vec{V}_2$ est pratiquement toujours égal au quart de la longueur d'onde λ des vibrations lumineuses.

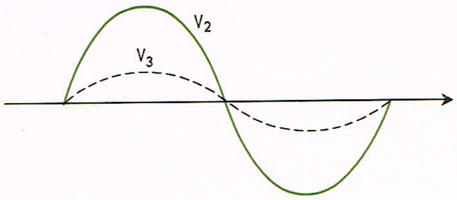
En résumé, la vibration $\vec{V}_1$ qui traverse A peut être remplacée par la somme :

- d'une vibration $\vec{V}_2$ qui est la même que dans le reste du champ (elle correspond à la lumière directe);
- d'une vibration $\vec{V}_3$ diffractée par A et dont l'amplitude b est d'autant plus faible que les indices de réfraction n et n' sont plus voisins (elle correspond à la lumière diffractée).

Dans le plan focal de l'objectif O, la vibration $\vec{V}_2$ (vibration directe) correspond à l'image S′ de la source S. La vibration $\vec{V}_3$ (vibration diffractée) correspond à la lumière qui s'étale autour de S′, et elle est décalée d'un quart de la longueur d'onde (λ/4) par rapport à $\vec{V}_2$.

Pour construire l'image P′ de la lame P donnée par l'objectif O, représentons les vibrations en A′ (image de A) et en B′ (image d'un point B quelconque du reste de la lame). Dans l'image A′ de A, on a les vibrations $\vec{V}_2$ et $\vec{V}_3$; à côté de A′, en un point B′ quelconque, la lumière est due à la seule vibration $\vec{V}_2$. Quelle est l'intensité lumineuse en A′, l'image de A ?

On voit que la vibration $\vec{V}_2$ est maximale quand la vibration $\vec{V}_3$ est minimale, et inversement. Leurs effets se séparent. A l'amplitude a de la vibration $\vec{V}_2$ correspond, comme nous l'avons dit, une intensité lumineuse égale à a^2; à l'amplitude b de la vibration $\vec{V}_3$, une intensité lumineuse égale à b^2. La théorie et l'expérience montrent que l'intensité lumineuse en A′ est égale à $a^2 + b^2$. Or, si le décalage d est petit, b est également petit et la quantité b^2 encore plus petite. On voit, par



exemple pour $a = 1$ et $b = 0{,}01$, que b^2, qui est égal à 0,000 1, est complètement négligeable devant a^2 égal à 1 : l'intensité lumineuse dans l'image A′ est donc pratiquement égale à a^2. En un point B′, à côté de A′, la lumière est formée par les vibrations $\vec{V}_2$; l'intensité est donc la même en tous les points de l'image P′, et la région A est invisible.

Supposons qu'il soit possible de décaler la vibration $\vec{V}_2$ d'un quart de longueur d'onde (nous verrons plus loin comment réaliser ce décalage). Les deux vibrations $\vec{V}_2$ et $\vec{V}_3$ s'ajoutent alors; l'amplitude résultante est égale à $a + b$. L'intensité lumineuse dans l'image A′ de la région A est égale à $(a + b)^2$, soit $a^2 + b^2 + 2ab$. Or, nous avons déjà dit que b^2 est négligeable devant a^2; l'intensité lumineuse en A′ est donc pratiquement égale à $a^2 + 2ab$ (on peut négliger b^2, mais non $2ab$). Nous constatons maintenant que l'intensité $a^2 + 2ab$ de l'image A′ est différente de l'intensité a^2 du reste du champ : la région A devient visible. **Le seul fait d'avoir décalé d'un quart de longueur d'onde les vibrations V_2 par rapport aux vibrations V_3 rend visible un objet transparent.** Décaler une vibration par rapport à une autre, comme on vient de le faire, s'appelle « déphaser » une vibration. Or, nous avons vu que l'objet transparent A décale la vibration V_1 qui le traverse par rapport à la vibration V_2 qui passe à côté; on dit, pour cette raison, que l'objet transparent est un objet « déphasant » ou objet « de phase ». On peut obtenir ainsi, à partir d'objets de phase, des images visibles possédant un certain contraste. D'où le nom de « contraste de phase » donné à la méthode.

la lame de phase, secret de la méthode

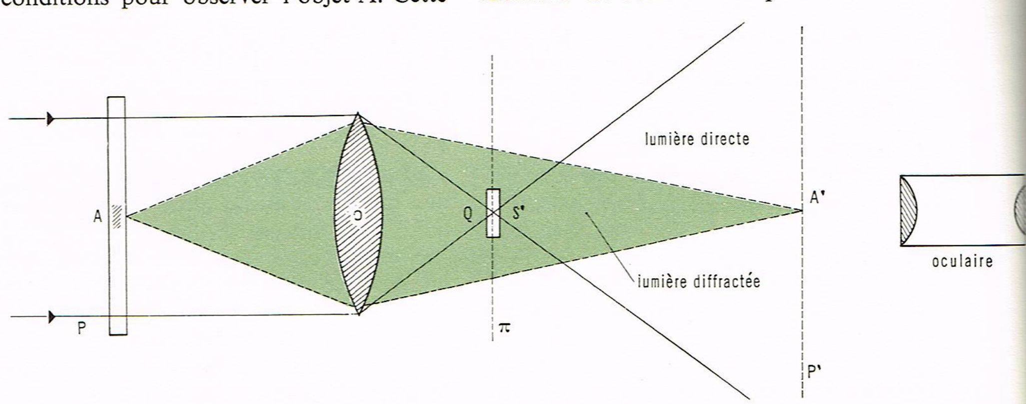
Il reste à savoir comment on peut retarder pratiquement les vibrations V_2, c'est-à-dire la lumière directe. Cette opération devient facile à comprendre si l'on veut bien considérer le schéma ci-dessous. Toute la lumière directe passe par S′, alors que la lumière diffractée s'étale largement dans cette région du plan π. Plaçons en S′ une petite lame transparente Q, dite « lame de phase ». Si elle est suffisamment étroite, on peut dire que, pratiquement, seul le faisceau de lumière directe (vibrations $\vec{V}_2$) la traverse, la fraction de lumière diffractée traversant la lame de phase pouvant être négligée. Il suffit maintenant de choisir l'épaisseur de la lame de phase Q, de telle manière qu'elle retarde exactement d'un quart de longueur d'onde les vibrations lumineuses qui la traversent.

On arrive très facilement à réaliser des lames de phase voulue par évaporation dans le vide de substances convenables. Si l'on retarde les vibrations directes, on dit que le contraste de phase est « négatif » : la région A est vue plus brillante que le reste du champ si $n' > n$. On peut aussi avancer les vibrations de la lumière directe d'un quart de longueur d'onde sans changer les vibrations de la lumière diffractée, et l'on dit alors que le contraste de phase est « positif » : la région A est vue moins brillante que le reste du champ si $n' > n$.

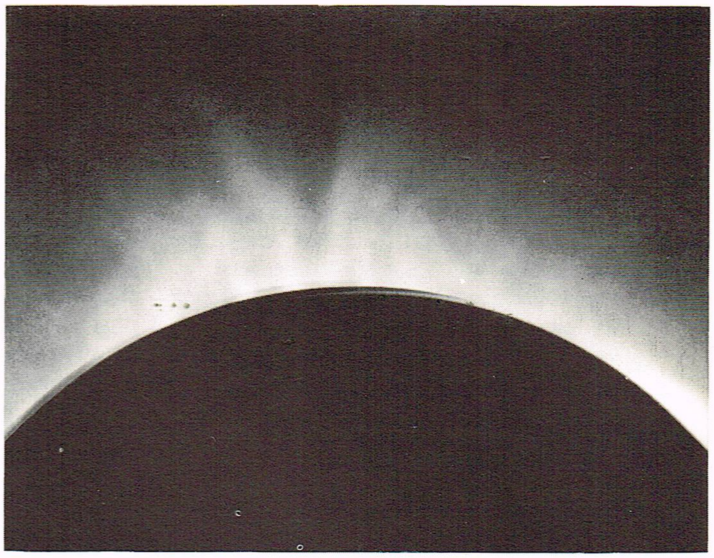
Si b est petit devant a (c'est-à-dire si n' est très peu différent de n), le contraste de l'image A′ n'est pas très grand, la valeur $a + b$ n'étant pas très différente de a. On peut améliorer considérablement le contraste des images en opérant de la façon suivante : au lieu de prendre une lame de phase transparente telle qu'une petite lame de verre, prenons une lame de phase absorbante. Cette modification ne change rien à la lumière diffractée (vibration $\vec{V}_3$), qui passe presque entièrement à côté de la lame de phase. En revanche, l'amplitude a des vibrations $\vec{V}_2$ (lumière directe) diminue d'autant plus que l'absorption de la lame de phase est plus grande. L'importance de b devant a augmente. La différence entre les intensités lumineuses $(a + b)^2$ et a^2 (en A′ et à côté) est d'autant plus grande que a est plus réduit par l'absorption de la lame en phase. On peut augmenter à volonté le contraste de l'image, et, par conséquent, sa sensibilité, en augmentant l'absorption de la lame de phase.

Il vient immédiatement à l'idée de pousser le raisonnement à la limite et de prendre une lame de phase « opaque ». Toute la lumière directe étant arrêtée, seule la lumière diffractée parvient à l'image A′, et, par conséquent, on doit être théoriquement dans les meilleures conditions pour observer l'objet A. Cette méthode bien connue constitue la méthode d'observation « en fond noir ». Cependant, l'expérience montre que la qualité des images est loin d'être celle qui est prévue par la théorie. En effet, puisqu'on arrête toute la lumière directe ($a = 0$), il ne reste que la lumière diffractée (vibrations d'amplitude b), dont l'intensité est proportionnelle à b^2. Mais il n'est plus possible maintenant de négliger b^2, puisqu'il n'y a pas d'autre terme. Or, si b est petit, b^2 est beaucoup plus petit, et l'on ne voit généralement pas d'image, car les sources lumineuses ne sont pas assez intenses. Même avec une source assez puissante, les images sont souvent masquées par la lumière diffusée par les imperfections des systèmes optiques.

Le fond noir est néanmoins utilisé en microscopie, car les petits objets que l'on observe correspondent à des vibrations dont l'amplitude b n'est jamais petite. C'est seulement dans le cas d'objets très faiblement déphasants que le fond noir donne de mauvais résultats, et c'est alors qu'il est indiqué d'appliquer la méthode du contraste de phase.



la réalisation du coronographe

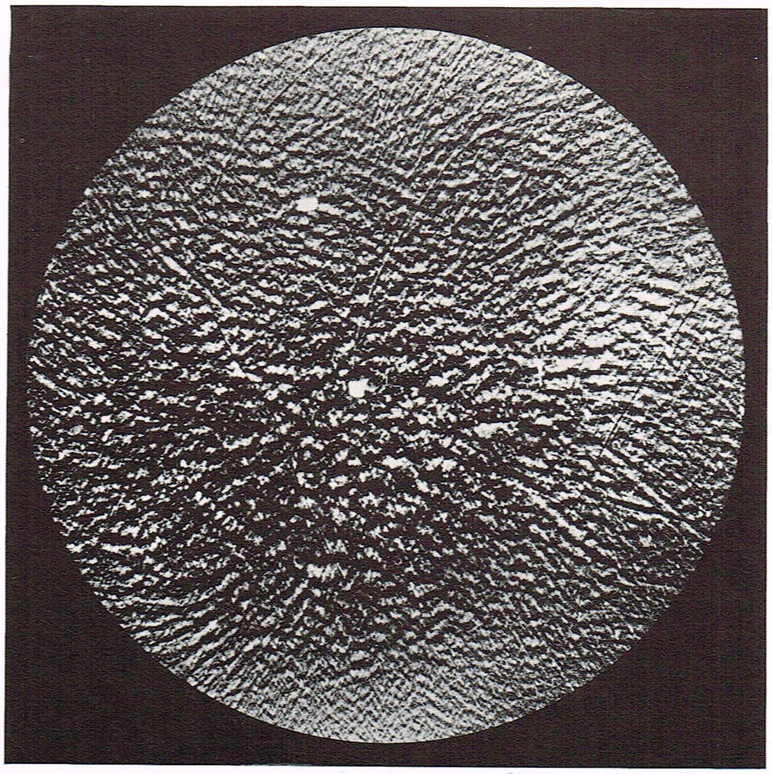


La première photographie monochromatique de la couronne obtenue sans éclipse, le 10 août 1939, à 9 h 20, au bord ouest, avec la raie verte, par Bernard Lyot.

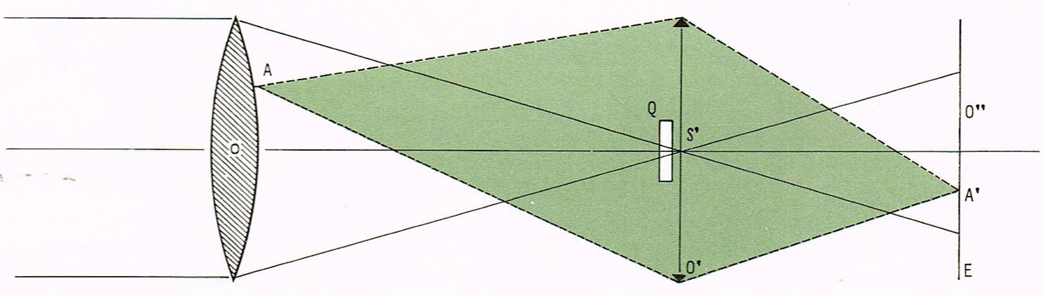
Revenons aux techniques du contraste de phase en employant une lame de phase absorbante, mais non opaque. Pour donner une idée de la sensibilité considérable ainsi obtenue, nous allons choisir un exemple d'application dans l'astronomie. On sait que la couronne solaire, parfaitement visible pendant une éclipse de Soleil, est très peu lumineuse par rapport à la photosphère (c'est-à-dire le Soleil tel que nous le voyons tous les jours), un million de fois plus lumineuse que la couronne. C'est parce que la photosphère est masquée par la Lune pendant une éclipse qu'il devient possible de voir la couronne. Les astronomes se sont efforcés d'observer la couronne en dehors des éclipses en masquant avec un petit écran opaque l'image du disque solaire donnée par un objectif. Mais leurs tentatives n'ont jamais abouti jusqu'en 1939 du fait de la lumière diffusée par les imperfections de l'objectif. En effet, l'objectif astronomique utilisé pour former l'image du Soleil en son foyer est éclairé par la photosphère, et toutes ses imperfections — défauts d'homogénéité (variations d'indice dans la masse du verre) et variations d'épaisseur (défauts de poli) — se traduisent par de la lumière diffractée. On a beau masquer l'image de la photosphère par un écran opaque, l'image de la couronne qui l'entoure se perd complètement dans la lumière diffusée, des milliers de fois plus importante.

Il fallait donc réaliser avant tout un objectif parfait et, pour cela, employer une méthode de contrôle extraordinairement sensible. L'astronome Bernard Lyot a réussi le premier, en 1939, à obtenir ce résultat remarquable en utilisant la méthode du contraste de phase. Pour contrôler l'objectif, on s'en sert à la fois comme objet et comme système optique formant l'image de la source. La lumière diffractée par un petit défaut A est recueillie par un objectif auxiliaire O′, qui forme une image de O sur une plaque photographique E. La lumière directe qui éclaire O traverse la lame de phase Q placée sur l'image S′ de la source (montage tout à fait analogue à celui qui a été décrit précédemment).

Le contrôle ainsi effectué est de très haute précision, et il a permis à Bernard Lyot de construire le premier coronographe et d'obtenir de magnifiques photographies de la couronne solaire en dehors des éclipses.



L'aspect d'une surface bien polie observée par contraste de phase lorsque la lame de phase a une forte absorption (grande sensibilité). Les dénivellations produites par les défauts de poli s'observent parfaitement et l'image est un peu analogue à la surface d'une mer agitée vue d'avion. Mais, ici, les vagues correspondent à des variations d'épaisseur de 0,000 002 millimètre, soit 2 millièmes de micron seulement !



microscopes à contraste de phase et à contraste interférentiel

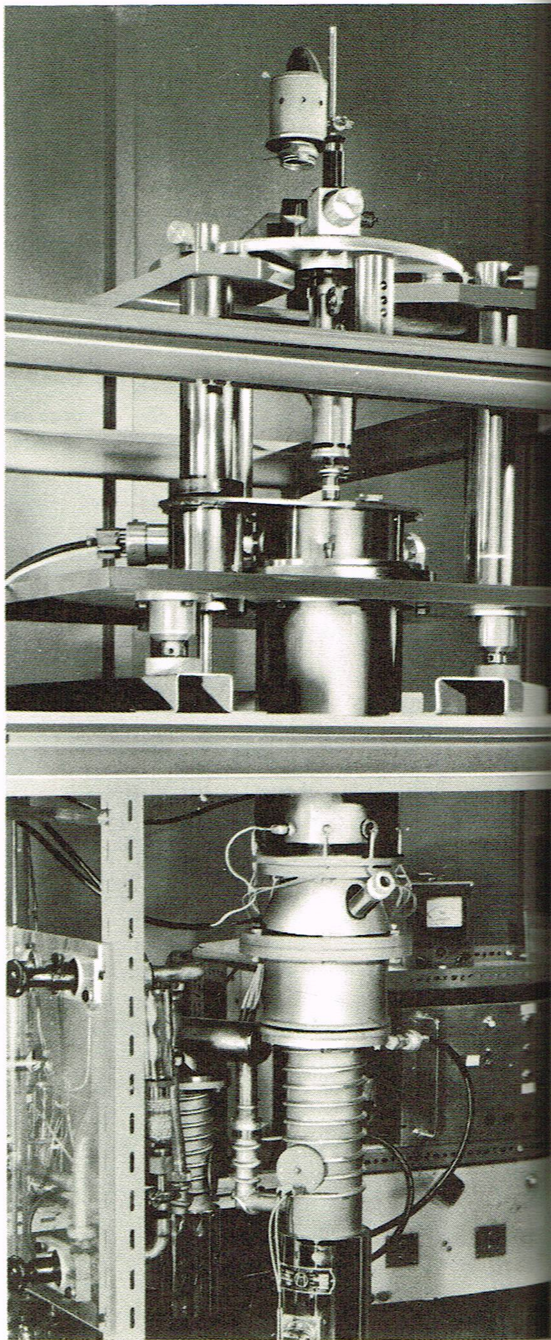
Si la méthode du contraste de phase s'applique à des objets de dimensions quelconques, c'est principalement en microscopie que ses applications sont les plus courantes. On devine son importance en biologie, puisque le microscope à contraste de phase permet l'observation des tissus vivants dans des conditions aussi avantageuses que celles auxquelles on est habitué dans l'observation des tissus morts après coloration. L'appareil conserve pratiquement l'aspect extérieur et le système optique des instruments habituels; il s'y ajoute seulement deux éléments, fort simples d'ailleurs, dont le fonctionnement se déduit immédiatement de ce qui précède. La figure peut être considérée comme donnant le schéma de principe du microscope ordinaire. Le condenseur est en C, la préparation en P, l'objectif en O et l'image P′ de P est observée par l'oculaire O*c*. Le seul élément à introduire pour transformer ce microscope ordinaire en microscope à contraste de phase est la lame de phase Q, qu'il faut placer en S′ au foyer de l'objectif du microscope. En fait, le schéma réel est un peu plus complexe; pour avoir plus de lumière, on remplace la source ponctuelle par une source annulaire, et l'on forme l'image d'une source étendue de façon qu'elle couvre complètement le diaphragme D percé d'un anneau (c'est cet anneau qui joue le rôle de source lumineuse). On a ici représenté le trajet des rayons de lumière directe issus d'un seul point de l'anneau. Sous l'action de la lumière directe, le détail transparent A diffracte la lumière, qui est recueillie par l'objectif O et concentrée dans l'image A′. La lame de phase Q est placée dans le plan focal de l'objectif O. C'est, par exemple, si l'on veut retarder les vibrations de la lumière directe, une lame de verre à faces parallèles sur laquelle on a déposé un petit anneau d'une substance transparente formant surépaisseur. Il faut naturellement que l'image de l'anneau D à travers C et O se fasse sur la lame de phase annulaire et que la lame de phase déborde un peu l'image de l'anneau D. Dans ces conditions, toute la lumière directe issue de l'anneau D passe dans la lame de phase, et le phénomène décrit précédemment a lieu. L'image A′ observée à l'aide de l'oculaire est une image en contraste de phase.

La technique de contraste de phase permet d'observer avec une extrême sensibilité des objets transparents. Mais les observations quantitatives, les mesures, sont difficiles, particulièrement en microscopie. C'est pourquoi les physiciens ont cherché à mettre au point des méthodes quantitatives, quitte à perdre, dans certains cas, un peu de sensibilité. Ce sont les interférences, et plus spécialement les interférences en lumière polarisée, qui ont permis les progrès les plus importants.

Expliquons ces phénomènes en observant les colorations bien connues des bulles de savon (*fig.* 2). Que se passe-t-il ? Un rayon lumineux SI, provenant d'une source lumineuse quelconque, se dédouble sur la bulle de savon B. Une partie du rayon incident SI se réfléchit en I sur la face extérieure de la bulle et l'autre partie se réfléchit en J sur la face intérieure. Bien entendu, la plus grande partie de l'énergie se trouve dans le rayon transmis JT, mais ce rayon ne nous intéresse pas, car nous observons seulement les rayons réfléchis. Comme l'absorption de la bulle de savon est négligeable, on peut dire que les deux rayons réfléchis ont la même énergie : ils transportent des vibrations dont les amplitudes sont égales. Il est facile de voir sur la figure qu'à partir de I les rayons n'accomplissent pas exactement le même trajet : le second rayon pénètre dans l'épaisseur de la bulle et accomplit un trajet un peu plus long que le premier rayon. Les vibrations du second rayon sont un peu en retard sur les vibrations transportées par le premier. Lorsque l'œil regarde la bulle, les vibrations transportées par les deux rayons se superposent sur la rétine de l'observateur. Ces deux rayons « interfèrent ». L'intensité lumineuse dans la région IJ dépend du retard du rayon 2 par rapport au rayon 1 de la figure 2.

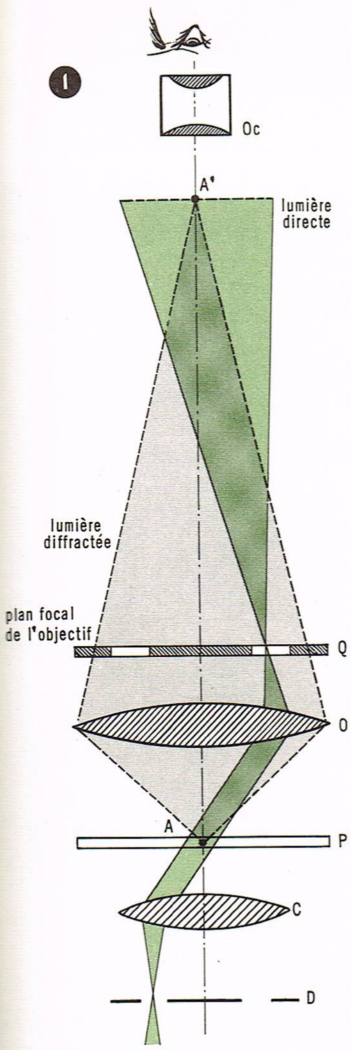
Naturellement, le phénomène n'est pas le même pour les différents points de la bulle, car son épaisseur et son indice de réfraction peuvent varier d'un point à un autre. Mais si l'épaisseur ou l'indice de la bulle varient, le décalage des vibrations n'est plus le même. L'intensité varie d'une région à l'autre de la bulle. Ces phénomènes, appelés « phénomènes d'interférences », expliquent les belles colorations des bulles de savon.

Un appareil interférentiel peut être fondé sur le principe décrit ci-dessus. Un rayon incident SI (*fig.* 4) s'est dédoublé en I par un dispositif dont nous ne préciserons pas la nature maintenant. Un premier rayon traverse l'objet dans la région où se trouve

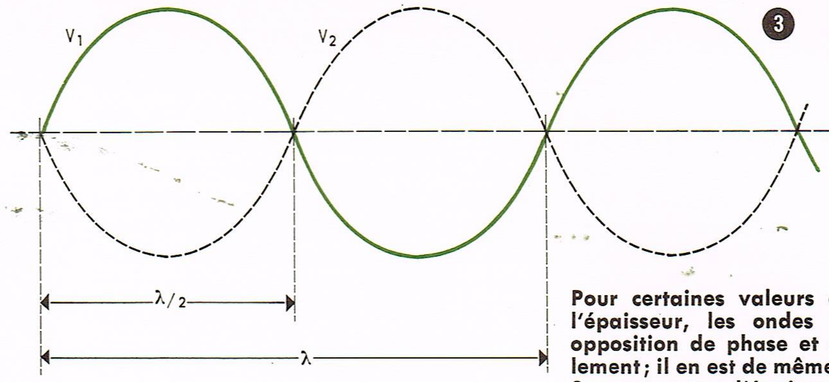


Microscope interférentiel pour les recherches atomiques.

❶



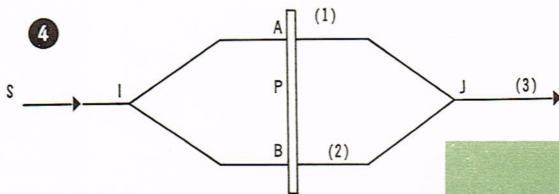
Microscope à contraste de phase.

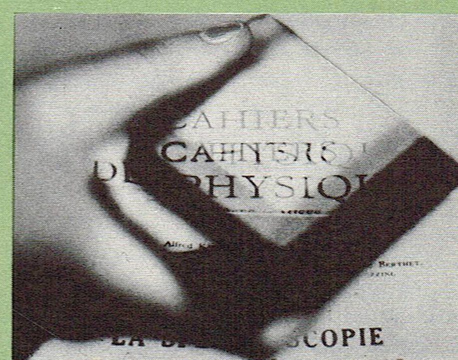


❸

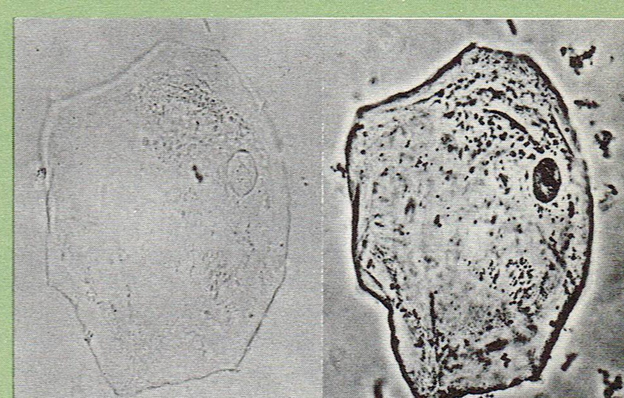
Pour certaines valeurs de l'indice et de l'épaisseur, les ondes peuvent être en opposition de phase et s'annuler mutuellement; il en est de même de l'éclairement. Supposons que l'épaisseur de la bulle soit précisément telle que le jaune ne soit pas réfléchi; la lumière blanche étant une superposition des couleurs du spectre, la lumière finale ne comprendra pas de composante jaune : la bulle apparaîtra pourpre.

❹





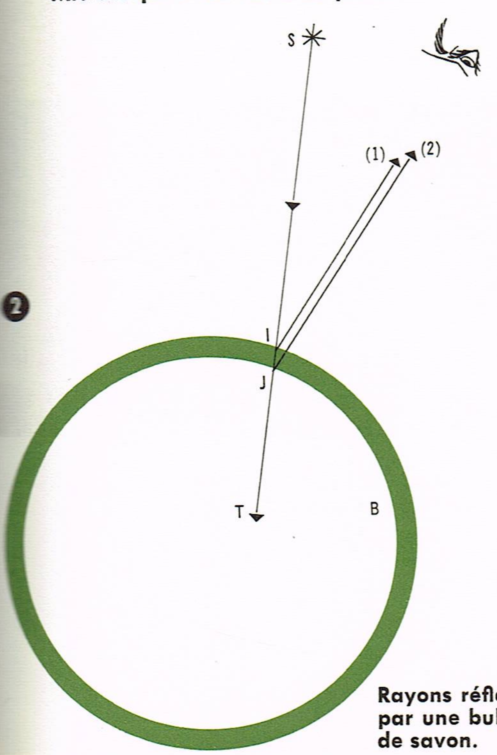
Dédoublement d'un objet vu à travers une lame de spath.



Cellule épithéliale de la langue vue au microscope ordinaire et en contraste de phase (grossissement 400).

Phot. Manigault. Institut Pasteur.

❷



Rayons réfléchis par une bulle de savon.

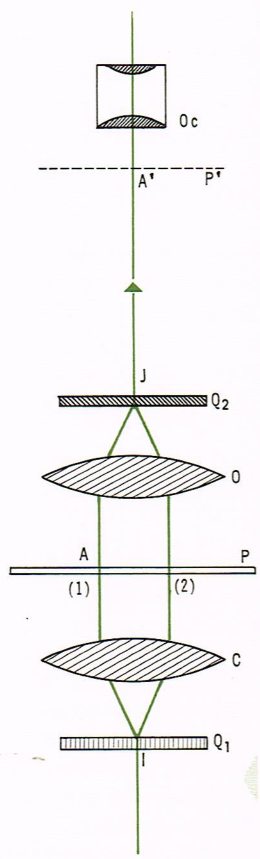


Schéma de principe d'un microscope à contraste interférentiel. Les rayons provenant de la source sont dédoublés par le système Q_1 (cristal de spath), ils traversent le condenseur C, la préparation P, l'objectif O et sont recombinés par le système Q_2 (cristal de spath). L'image P' de P donnée par l'objectif O est observée par l'oculaire Oc. L'image P' est une image en contraste interférentiel.

Phot. Manigault, Institut Pasteur.

le détail A à observer. Un second rayon passe dans une autre région B de P où il n'y a pas de détail à étudier. Ces deux rayons sont de nouveau confondus à partir du point J grâce à un dispositif analogue à celui qui se trouve en I. Par construction, les deux trajets ont une différence bien déterminée si la lame P a une épaisseur et un indice parfaitement constants (pas de variation d'indice en A).

Si l'objet P possède en A un indice n différent de l'indice n' du reste de la lame (avec $n>n'$, par exemple), les vibrations transportées par le premier rayon sont retardées, et le décalage des vibrations transportées par les deux rayons n'est plus le même. Ce décalage reste inchangé dans la région où les rayons sont confondus, et jusque sur la rétine de l'observateur. On peut alors reprendre le raisonnement qui a permis d'expliquer les colorations des bulles (*fig.* 3, p. 161). Supposons que la différence initiale des trajets ($n \neq n'$) soit telle que les vibrations transportées soient en opposition à la sortie de l'appareil pour une longueur d'onde jaune. Les vibrations se détruisent et la lame est vue colorée en pourpre, alors que S émet de la lumière blanche. Si, maintenant, il y a en A une région d'indice n différent du reste de la lame, A est vu sous une coloration différente de celle du reste de l'objet : on a décelé l'existence d'un objet pourtant transparent. L'objet A est vu en contraste interférentiel. Comme la couleur en A est caractéristique de la différence, **il s'agit d'une méthode quantitative.**

La polarisation des rayons a fait qualifier ces appareils de « contraste interférentiel à polarisation », et c'est dans ce domaine que les progrès les plus importants ont été réalisés.

Si le détail A est assez large, il faut un dédoublement (distance séparant les deux rayons) important, sinon les images risquent — notamment en microscopie — de s'enchevêtrer inextricablement. Pour éviter ce fâcheux inconvénient, on peut donner au dédoublement une valeur assez petite pour qu'il soit invisible dans le système d'observation. Les images observées ont un aspect différent de celui des images précédentes et la sensibilité est un peu moins grande. Mais la méthode peut être employée dans n'importe quel cas. Le spath n'est pas le seul cristal à posséder la propriété de dédoubler les rayons; on emploie de préférence le quartz, plus facile à trouver et de meilleure qualité au point de vue optique. Sauf pour certaines applications spéciales, le microscope à contraste interférentiel n'a pas un aspect différent de celui du microscope ordinaire.

couches minces et phénomènes d'interférence

Les bulles de savon dont nous avons parlé précédemment fournissent un exemple de couche mince. Lorsque l'épaisseur de la bulle est de l'ordre de quelques dixièmes de micron, de vives couleurs apparaissent lorsqu'on regarde la lumière qu'elle réfléchit. Ces couleurs sont dues aux interférences des rayons réfléchis par les deux faces de la bulle, et nous avons étudié le mécanisme de production de ces couleurs. Les physiciens désignent par le nom de « couches minces » des lames très fines à faces parallèles, qui produisent exactement les mêmes effets d'interférence, mais sont réalisées avec une très grande précision. On sait déposer des couches minces très résistantes sur des surfaces et prévoir leur épaisseur. Leurs applications sont innombrables : traitement antiréfléchissant des surfaces, filtres interférentiels, filtres anticalorifiques, miroirs froids...

comment réaliser une couche mince

Le problème que l'on se pose est le suivant : on veut déposer sur une surface de verre d'indice N une couche mince d'un produit transparent d'indice n. L'épaisseur e peut varier suivant les cas; fixons un ordre de grandeur en disant que l'épaisseur sera par exemple de 0,1 μ (0,000 1 mm). Le procédé le plus employé est celui de l'évaporation sous vide.

Une cloche, en métal ou en verre, est posée sur un plateau métallique, l'ensemble formant une enceinte absolument étanche. Le plateau est percé en son centre, et un tuyau met en communication l'enceinte et une pompe à faire le vide, non représentée sur la figure. La lame de verre sur laquelle on veut déposer la couche mince est en M, sur un support.

La substance à évaporer est placée en contact avec un filament métallique relié à un générateur de courant électrique (non représenté sur la figure) par l'intermédiaire des fils A et B. La pression de l'air dans la cloche est contrôlée par des organes électriques. Lorsque la pression est de l'ordre d'un millionième de millimètre de mercure, celle d'un vide déjà très poussé, on fait passer un courant électrique, qui chauffe le filament F et, du même coup, la substance à évaporer. Lorsque la substance atteint sa température de fusion, elle s'évapore et se dépose partout dans la cloche, et en particulier sur la lame de verre M. Naturellement, il ne faut pas que le filament s'évapore lui aussi; il suffit pour cela que la température de fusion du filament soit plus élevée que celle de la substance. Des appareils de contrôle très précis permettent de connaître l'épaisseur de la couche pendant l'évaporation et d'arrêter celle-ci lorsque l'épaisseur a atteint la valeur désirée.



Appareil pour déposer des couches min[...] sous vide et, *ci-contre*, schéma de princ[...]

Phot. Zeiss.

Il est même possible de déposer par cette méthode, les unes sur les autres, plusieurs couches dont les épaisseurs et les indices sont connus. Un tel empilement de couches a des propriétés très intéressantes, comme nous le verrons plus loin. Citons à titre d'exemple, parmi les substances utilisées pour réaliser des couches minces transparentes : le fluorure de magnésium et le sulfure de zinc.

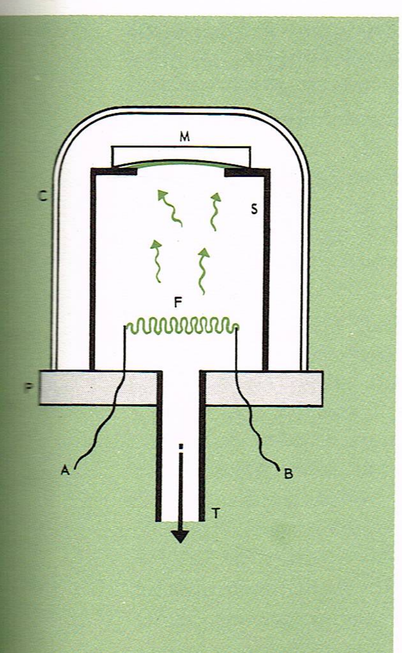
une lumière qui ne peut plus se réfléchir

Imaginons une couche mince d'épaisseur e et d'indice n déposée sur une surface de verre par le procédé précédent. Comparons au schéma de la bulle d'air ; il y avait de l'air des deux côtés de la bulle, et maintenant il y a de l'air d'un côté de la couche mince et du verre de l'autre côté. Les conditions physiques ne sont pas tout à fait les mêmes dans les deux expériences, mais le principe fondamental des phénomènes est le même. Les rayons réfléchis par les deux faces de la couche mince transportent des vibrations comme dans le cas des rayons réfléchis par les faces de la bulle. Pour une épaisseur e et un indice n convenables de la lame, il peut très bien se faire que les deux vibrations soient encore en opposition pour une longueur d'onde déterminée : pour cette longueur d'onde il n'y a pas de lumière réfléchie. Le calcul et l'expérience montrent qu'il en est ainsi pour une longueur d'onde λ telle que $\boldsymbol{ne = \lambda/4}$, c'est-à-dire que le seul fait de déposer sur une surface de verre d'indice N une couche mince transparente d'épaisseur e telle que $ne = \lambda/4$ (lame « quart-d'onde », avec $n < N$) supprime la lumière réfléchie par la surface de verre. Puisque la lumière ne peut pas se réfléchir, elle traverse la couche mince et pénètre dans le verre. Si l'on dépose sur la surface inférieure de la lame de verre une couche mince identique à la précédente, la lumière ne peut pas se réfléchir non plus sur elle et traverse la lame. Par conséquent, pour une longueur d'onde donnée, tout le flux lumineux incident se retrouve à la sortie de la lame de verre et il n'y a aucune perte d'énergie par réflexion sur ses faces. Evidemment, il est impossible d'annuler la réflexion pour toutes les radiations en même temps ; le phénomène se produit pour une longueur d'onde et non pour les autres. Il est naturel de choisir la longueur d'onde, c'est-à-dire la couleur, pour laquelle l'œil possède la sensibilité maximale, de manière à réduire au mieux les pertes de lumière par réflexion. Cette couleur se trouve au voisinage du jaune ; il faut donc choisir une épaisseur e et un indice n tels que cette couleur soit précisément éteinte par réflexion : en lumière blanche on observe alors une couleur pourpre foncé peu visible. C'est bien ce que l'on cherche : diminuer le plus possible la lumière réfléchie (l'idéal serait certes le noir, mais nous avons déjà vu que cela n'est pas possible). La lame de verre ainsi recouverte de deux couches sur ses deux faces est dite « traitée ». Si l'on observe un objet quelconque à travers cette lame, très peu de lumière est perdue par réflexion sur ses faces : sa transparence est maximale.

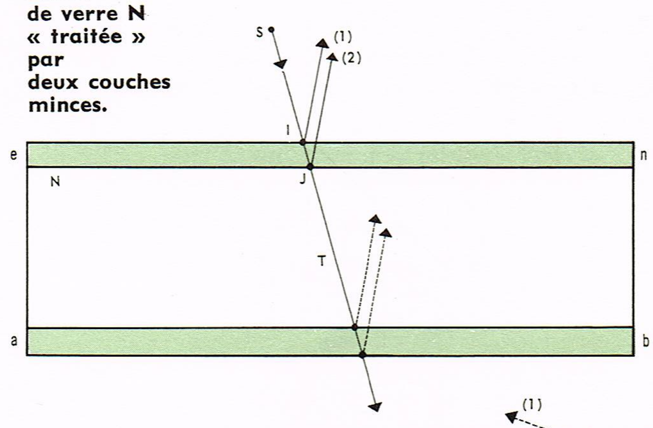
Prenons par exemple une lame de verre d'indice $N = 1{,}52$, recouverte par une couche mince de fluorure de magnésium d'épaisseur e et d'indice n tels que $ne = \lambda/4$ pour $\lambda = 0{,}55\ \mu$ (jaune-vert). La lumière réfléchie est quatre fois plus faible que pour le verre nu.

pourquoi « traiter » les instruments ?

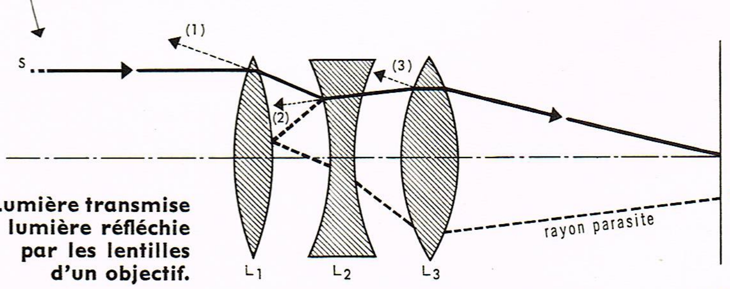
Prenons pour l'exemple un objectif photographique formé par trois lentilles. En arrivant sur la lentille L_1, une petite partie du rayon incident se réfléchit sur les deux faces de L_1 ; il en est de même lorsque le rayon arrive sur L_2 et sur L_3. On a représenté seulement les rayons réfléchis par les faces d'entrée de ces lentilles. Le rayon en gros trait représente le rayon normal qui aboutit en P sur l'image. Le rayon réfléchi sur la première face de L_1 est définitivement perdu, mais, pour les autres, il n'en est pas de même. En repartant vers la gauche, ces rayons sortent de l'objectif, mais une petite partie de chacun



Lame de verre N « traitée » par deux couches minces.



UNE LUMIÈRE QUI NE PEUT PLUS SE RÉFLÉCHIR



Lumière transmise et lumière réfléchie par les lentilles d'un objectif.

d'eux est réfléchie à nouveau par les lentilles et repart vers la droite. Bien entendu, cette petite fraction de lumière ne concourt pas à la formation de l'image; c'est de la lumière parasite, qui voile l'image en P. Par conséquent, non seulement la lumière réfléchie par les faces des lentilles est perdue quant à la formation de l'image, mais — ce qui est plus grave — elle produit un voile parasite sur l'image. On devine comment remédier à ces inconvénients par l'emploi des couches minces. Déposons sur toutes les faces des lentilles des couches minces identiques aux couches minces déposées sur la lame de verre. Les pertes par réflexion sont réduites considérablement et, du même coup, la transparence de l'objectif est augmentée. **Comme le « traitement » a pour effet de faire disparaître presque complètement le voile parasite, on obtient donc un gain sur la luminosité et sur la qualité de l'image.** Les résultats sont d'autant plus intéressants que l'instrument comporte un plus grand nombre de lentilles. Par exemple, dans un périscope de marine ou dans un cystoscope, il peut y avoir jusqu'à trente surfaces air-verre, et, dans ce cas, 80 p. 100 de la lumière incidente sont perdus par réflexion sur ces surfaces. On conçoit, dans ces conditions, l'intérêt qu'il y a à « traiter » les surfaces des lentilles. Avec une seule couche mince, il est possible d'annuler la réflexion pour une longueur d'onde, comme nous venons de le voir. On peut améliorer encore les résultats en remplaçant une couche mince par deux ou même trois couches minces. En donnant à ces couches des épaisseurs et des indices convenables, la réflexion est annulée pour trois longueurs d'onde dans le cas de trois couches. Donnons un exemple : dans un instrument d'optique qui comporte vingt surfaces air-verre non traitées, 32 p. 100 seulement du flux incident parviennent à l'image. Après traitement à l'aide de trois couches minces, 98 p. 100 de la lumière sont utilisés.

les couches minces peuvent aussi servir à filtrer la lumière

Lorsque l'on utilise une couche mince déposée sur une lame de verre, on peut annuler la réflexion pour une longueur d'onde, par exemple le jaune, et toutes les autres couleurs sont réfléchies. En combinant couches minces transparentes et couches métalliques très minces (elles réfléchissent, mais elles laissent passer aussi une certaine fraction de lumière), on sait réaliser le phénomène inverse : réfléchir une seule longueur d'onde (une seule couleur) et annuler la réflexion pour toutes les autres. L'ensemble constitue un « filtre interférentiel par réflexion ». Eclairé en lumière blanche, un tel filtre ne réfléchit qu'une seule couleur. Ce qui est réalisable par réflexion l'est également par transmission : en déposant sur une lame de verre un ensemble de couches minces transparentes et métalliques, on peut arrêter toutes les couleurs, sauf une. Le dispositif constitue un « filtre interférentiel par transmission ». Les filtres interférentiels permettent de produire une lumière de couleur déterminée, c'est-à-dire une lumière de longueur d'onde connue à l'avance et dont la pureté peut aussi être réglée suivant les besoins. Toutes les combinaisons sont possibles avec eux. On peut isoler une très petite portion du spectre et obtenir ainsi un filtre très « monochromatique » ou, au contraire, une région très large du spectre, position et largeur de la bande colorée transmise ou réfléchie pouvant être choisies à volonté. De tels filtres remplacent avantageusement les gélatines et les verres colorés, qui sont très peu monochromatiques et pour lesquels le choix des caractéristiques n'est pas possible.
Les filtres interférentiels sont utilisables non seulement dans le domaine des longueurs d'onde visibles, mais aussi en infrarouge et en ultraviolet. On réalise des filtres bien réfléchissants dans l'infrarouge et qui laissent passer le visible : ce sont des filtres anticaloriques. Inversement, il est intéressant d'avoir des miroirs qui réfléchissent bien le visible et laissent passer les radiations infrarouges sans les absorber. Ces « miroirs froids » sont obtenus aussi par l'emploi de couches minces superposées. Ils sont utiles dans les appareils de projection de haute intensité où l'élimination de la chaleur devient un problème important. Les couches minces sont même utilisées dans les satellites artificiels pour régulariser la température (satellites « Vanguard »). L'intérêt des couches minces ne réside pas seulement dans leurs applications, qui sont multiples, comme on vient de le voir; elles permettent aussi d'approfondir la conception que l'on se faisait de certains phénomènes : elles ont notamment été utilisées dans l'étude de la structure électronique des métaux et ont permis de vérifier des prévisions théoriques faites sur les métaux supraconducteurs.

de nouvelles sources lumineuses : les lasers

Des progrès considérables ont été réalisés dans le domaine des sources lumineuses. Grâce à une source à krypton, il est possible de mesurer directement en longueurs d'onde le mètre étalon en platine iridié du Bureau international des poids et mesures. Le prototype en platine est remplacé maintenant par une série de longueurs d'onde identiques mises bout à bout. Leur ensemble constitue une « longueur » qui est par définition celle du mètre étalon. Mais les sources les plus extraordinaires sont encore les « lasers » (« Light Amplification by Stimulated Emission of Radiation »). **Leur principe de fonctionnement est très différent de celui des autres sources. Par leur puissance et leur finesse, ils ouvrent une ère nouvelle non seulement en optique, mais dans toutes les branches de la science.**

mesure et définition du mètre étalon

Nous avons montré comment les phénomènes d'interférence permettent d'expliquer les colorations des bulles de savon : une très faible variation d'épaisseur provoque un changement de teinte. Imaginons que l'épaisseur de la bulle augmente.

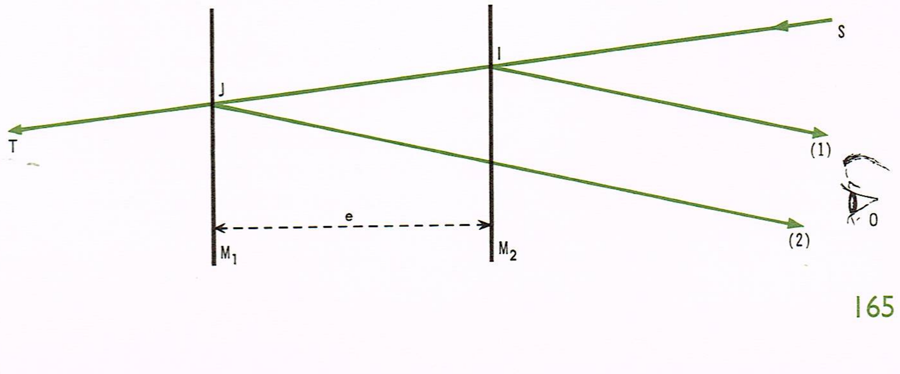
A partir d'un micron environ, les colorations disparaissent. Il n'est plus possible d'observer des interférences en lumière blanche, mais on peut continuer à les voir en lumière colorée. Eclairons la bulle avec une lampe ordinaire devant laquelle on a placé un écran transparent coloré en jaune, un filtre interférentiel par exemple. Supposons que l'épaisseur de la bulle soit variable sur toute sa surface. Si le filtre jaune ne laisse passer qu'une seule longueur d'onde, c'est-à-dire s'il est parfaitement monochromatique, les vibrations transportées par les rayons sont sinusoïdales. Dans une région d'épaisseur déterminée, il est possible que les vibrations transportées par les rayons réfléchis soient en opposition de phase : elles se détruisent, la bulle paraît alors noire. Dans une autre région, les vibrations peuvent au contraire additionner leurs effets, et la bulle paraîtra très brillante. La surface de la bulle entière semblera sillonnée de régions noires entre lesquelles se trouveront des régions brillantes. Les régions noires sont appelées « franges noires » et les régions brillantes « franges brillantes ».

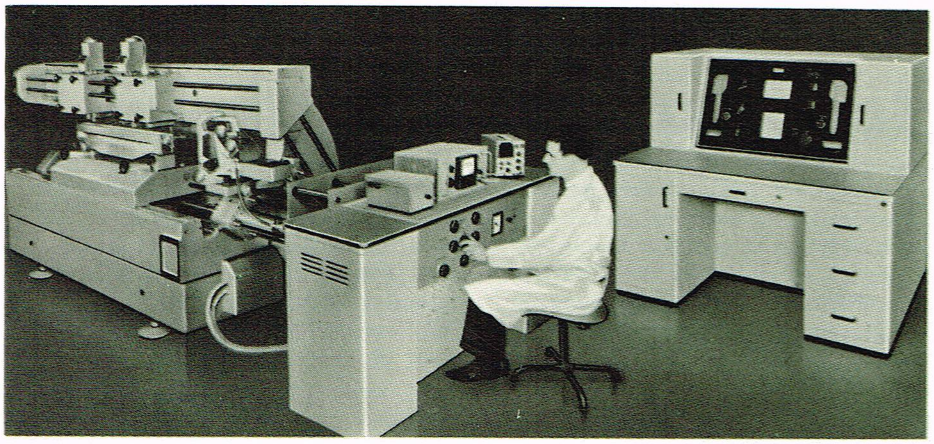
Partons d'un point situé sur une frange noire. Si nous considérons tous les points situés sur la même frange noire, nous savons que l'épaisseur de la bulle ne change pas. Mais si nous nous écartons de cette frange pour atteindre une région éclairée, nous savons que l'épaisseur change et que les vibrations glissent l'une par rapport à l'autre : sur une frange noire les deux vibrations $\vec{V}_1$ et $\vec{V}_2$ sont en opposition. Imaginons que l'épaisseur augmente toujours. Les deux vibrations $\vec{V}_1$ et $\vec{V}_2$ continuent à glisser l'une par rapport à l'autre et chaque fois qu'elles se retrouvent en opposition il n'y a pas de lumière réfléchie (frange noire sur la bulle). Cela veut dire que si la source émet des vibrations ayant une longueur d'onde déterminée (lumière parfaitement monochromatique), on peut voir des interférences quelle que soit l'épaisseur.

Remplaçons la bulle, dont l'épaisseur reste toujours petite, par des lames de verre d'épaisseur 1, 20, 50, 100 mm. On continue d'observer, très bien, en lumière monochromatique, les phénomènes d'interférence même pour des épaisseurs de plus en plus grandes. Pour obtenir des épaisseurs encore plus grandes, remplaçons la lame de verre par deux surfaces dont on peut faire varier la distance e à volonté. Evidemment, ces deux surfaces doivent appartenir à deux lames de verre, que nous supposerons, pour simplifier l'exposé, infiniment fines. Elles sont représentées par les deux traits parallèles M_1 et M_2. On comparera cette figure au schéma de la bulle de savon. Les deux surfaces M_1 et M_2 sont parfaitement parallèles et séparées par un intervalle d'air. Elles sont éclairées par un faisceau de rayons parallèles. Le rayon SI est un rayon du faisceau incident. Si les vibrations transportées par les deux rayons sont en opposition, cette disposition est la même pour tous les rayons réfléchis du faisceau, et l'ensemble M_1 M_2 ne réfléchit pas la lumière. C'est ce que constate l'œil de l'observateur placé en O.

Modifions de très peu l'écartement e des surfaces M_1 et M_2 : le phénomène change aussitôt. De la lumière réapparaît, mais, en continuant à faire varier e, nous retrouvons une position pour laquelle la lumière réfléchie est de nouveau supprimée. Partons de $e = 0$, et augmentons d'une façon continue la distance entre M_1 et M_2. Le phénomène se répète périodiquement : beaucoup de lumière, obscurité, beaucoup de lumière, obscurité, etc. : on voit défiler les franges. Si la lame a une épaisseur irrégulière, on voit les franges sur la lame; si la lame est à faces parallèles, on retrouve les franges noires et les franges brillantes en faisant varier la distance des faces. La théorie montre que l'on passe d'une frange brillante à la frange brillante suivante lorsque la distance M_1 M_2 augmente ou diminue de $\lambda/2$. Pour le jaune, on a $\lambda/2 = 0{,}3\ \mu$, une très petite valeur.

Il est maintenant facile de comprendre comment on peut utiliser les phénomènes d'interférence pour mesurer une longueur. La longueur à mesurer est représentée par une certaine distance e entre deux surfaces M_1 et M_2. Réduisons-la progressivement : on voit défiler les franges noires et les franges brillantes. Repérons les franges brillantes par exemple et comptons combien nous en voyons défiler





Étalonnage d'un mètre métallique avec l'étalon au krypton.

jusqu'au moment où les deux surfaces sont au contact ($e = 0$). Si nous voyons défiler 100 000 franges, l'épaisseur e, c'est-à-dire la longueur à mesurer, est égale à 100 000 $\lambda/2$, soit 30 mm. La mesure consiste donc à déterminer combien la longueur à étudier contient de demi-longueurs d'onde de la lumière utilisée. En principe, si la lumière est parfaitement monochromatique, on peut mesurer toute longueur, si grande soit-elle. Mais les sources lumineuses ne sont jamais rigoureusement monochromatiques et il n'est pas possible de mesurer n'importe quelle longueur. Avec de la lumière blanche on ne peut guère dépasser des épaisseurs supérieures au micron. Au fur et à mesure que la lumière devient plus monochromatique, on peut observer des franges avec des épaisseurs de plus en plus grandes.

Peut-on aller jusqu'au mètre, et mesurer ainsi le mètre en longueurs d'onde? Jusqu'à ces dernières années, cela n'était pas possible d'une manière directe; il fallait procéder par étapes. On connaît maintenant une source qui produit une lumière assez monochromatique pour comparer directement le mètre étalon en platine à une longueur d'onde. La source est une source à krypton, qui émet des vibrations de 0,605 6 μ de longueur d'onde (couleur orangée). **Le mètre est maintenant défini comme la longueur qui contient 1 650 763,73 longueurs d'onde des vibrations émises par la source précédente dans des conditions déterminées.**

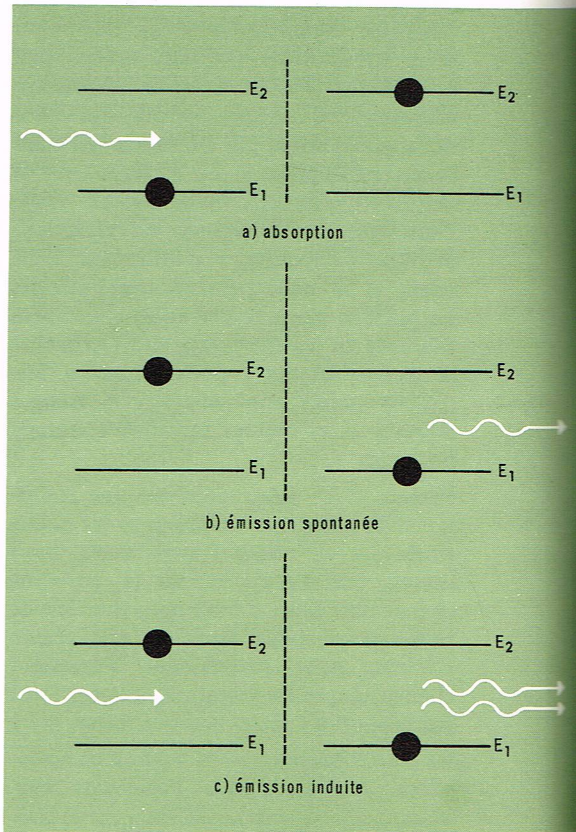
On a ainsi un étalon de longueur précis et sûr. Même si le mètre étalon du Bureau international des poids et mesures venait à disparaître pour une raison ou une autre, rien ne serait désormais perdu : les hommes retrouveraient leur unité fondamentale de longueur.

maser et laser

Lorsqu'un atome reçoit de l'énergie, il est porté à un état d'énergie supérieure, et c'est en revenant à l'état initial ou à un état intermédiaire qu'il émet un rayonnement. Cette énergie peut lui être fournie par des photons. Si l'atome absorbe un photon, il passe de l'état de plus basse énergie E_1, qui est l'état fondamental, à un état d'énergie supérieure E_2. On dit que l'atome est excité. En général, l'atome revient à l'état E_1 en émettant spontanément un photon : il y a émission spontanée de rayonnement. La longueur d'onde λ du rayonnement émis est donnée par la formule $E_2 - E_1 = \frac{hc}{\lambda}$, où h est une constante et c la vitesse de la lumière. Il peut arriver aussi que le retour ($E_2 \rightarrow E_1$) soit provoqué par un photon : on dit qu'il y a émission « induite » de rayonnement. Par l'émission induite, l'atome émet un photon et celui-ci vient s'ajouter au photon qui a provoqué l'émission. A ces photons correspondent des ondes et, fait remarquable, l'onde incidente qui a provoqué l'émission induite est en phase avec l'onde émise par l'atome. Nous insistons sur ce point, car c'est là que réside la distinction la plus importante entre les deux types d'émission. Dans l'émission induite, l'onde émise a la même phase que l'onde incidente, et l'on dit que l'onde incidente conserve sa « cohérence de phase » dans ce processus. L'amplitude de l'onde incidente est augmentée par l'onde émise par l'atome. **Le rayonnement produit par émission induite est « cohérent ».**

Au contraire, la production de rayonnement par émission spontanée est essentiellement un processus aléatoire, c'est-à-dire qu'il se fait au hasard. **C'est pourquoi l'émission spontanée n'a pas de cohérence de phase.**

Considérons maintenant une substance dans laquelle se trouvent un grand nombre d'atomes. D'après la théorie quantique, les atomes ne peuvent exister que dans des états ayant des énergies bien définies E_1, E_2, E_3, ... L'état d'énergie E_1 sera par exemple l'état fondamental, c'est-à-dire l'état de plus basse énergie, et les autres seront les états excités. Un certain nombre d'atomes se trouveront à un niveau d'énergie quelconque E_3, ces atomes



Phot. Bureau des poids et mesures.

représentant la population du niveau E_3. Plus un niveau est élevé dans l'échelle des niveaux et plus sa population est faible. Cette répartition des populations correspond bien entendu à un équilibre statistique, et les atomes vont continuellement subir des transitions d'un état à un autre. En absorbant des photons incidents, les atomes passent du niveau E_1 au niveau E_2, par exemple. Ils retournent au niveau E_1 soit par émission spontanée, soit par émission induite. L'émission induite joue un rôle d'autant plus grand que la longueur d'onde est plus grande, prédominant même dans le domaine des ondes hertziennes, où les longueurs d'onde sont de l'ordre du centimètre ou du mètre. Au contraire, dans le domaine optique, c'est l'émission spontanée qui est le phénomène prédominant. Plaçons-nous dans le cas où l'émission spontanée est négligeable et considérons seulement deux niveaux E_1 et E_2. Puisque $N_1 > N_2$, la probabilité est plus grande pour que les photons incidents fassent monter les atomes du niveau E_1 au niveau E_2 (absorption) plutôt qu'ils ne provoquent la descente du niveau E_2 au niveau E_1 (émission induite).

Il y a plus de photons absorbés que de photons émis : le rayonnement s'affaiblit. **Imaginons que, par un artifice quelconque, il soit possible d'inverser les populations des niveaux E_1 et E_2 et de faire $N_2 > N_1$. Que va-t-il se passer?** Cette fois, il y a plus de probabilité pour que les photons incidents provoquent la descente $E_2 \rightarrow E_1$ (émission induite) que la montée $E_1 \rightarrow E_2$ (absorption), car $N_2 > N_1$. L'émission induite va l'emporter sur l'absorption. Au lieu que des photons soient pris au rayonnement incident (absorption), d'autres lui seront ajoutés (émission induite). L'intensité du rayonnement incident va augmenter; il y a amplification. D'après les remarques précédentes au sujet de la phase, on peut dire qu'il y a amplification véritable et non un apport aléatoire. L'onde sinusoïdale incidente grandit en amplitude et conserve sa cohérence de phase.

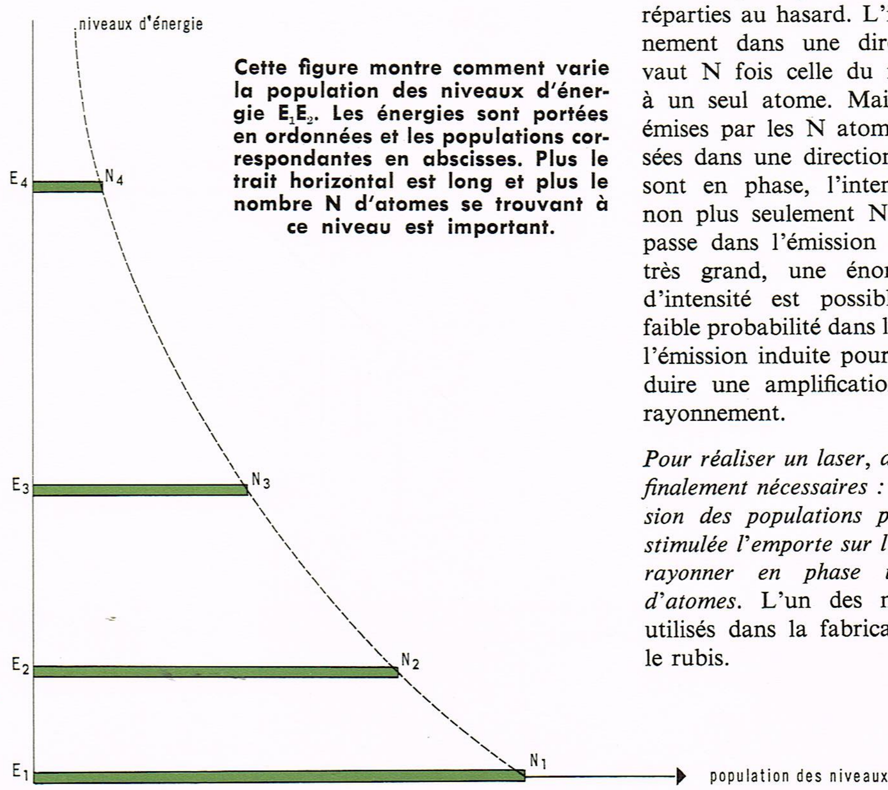
Donc, si un rayonnement pénètre dans un milieu où les populations des niveaux d'énergie E_1 et E_2 sont inversées, il y a amplification, et le rayonnement émis a une intensité considérablement supérieure à celle du rayonnement incident. C'est le principe des « masers » (de « Microwaves Amplification by Stimulated Emission of Radiation »). Le premier maser a été imaginé et réalisé en 1954 par Townes, à l'université Columbia aux Etats-Unis : l'inversion des populations a été obtenue en faisant passer un faisceau d'atomes se trouvant dans les états E_1 et E_2 à travers un filtre qui laisse passer les atomes dans l'état E_2 et rejette ceux qui sont dans l'état E_1. Pour que le système fonctionne, il faut maintenir en permanence ou pendant un certain temps un excès de population du niveau supérieur E_2. On a intérêt alors à réduire au minimum les pertes par rayonnement du système, en enfermant la matière (formée par les atomes dont nous avons parlé) dans une boîte dont les parois sont, en quelque sorte, « réfléchissantes ».

L'ammoniac, « matière » utilisée par Townes, présente deux états E_1 et E_2 pour lesquels la longueur d'onde émise est de l'ordre de 3 cm. Si l'on veut répéter l'expérience dans le domaine optique, il faut réaliser l'inversion de population pour deux niveaux E_1 et E_2 dont la différence d'énergie corresponde à une longueur d'onde optique, par exemple 0,6 μ (couleur jaune). Alors que l'émission spontanée pouvait être négligée précédemment, il n'en est plus de même maintenant, dans le domaine optique. C'est l'émission induite qui est faible lorsqu'elle est comparée à l'émission spontanée. Toutefois, grâce à sa cohérence de phase, l'émission induite peut, malgré sa faible probabilité, produire une intensité beaucoup plus grande que l'émission spontanée. Cela peut se comprendre de la manière suivante : si nous avons un grand nombre N d'atomes émettant spontanément, les phases des vibrations sont réparties au hasard. L'intensité du rayonnement dans une direction quelconque vaut N fois celle du rayonnement I dû à un seul atome. Mais si les vibrations émises par les N atomes sont synchronisées dans une direction où les vibrations sont en phase, l'intensité sera N^2I, et non plus seulement NI. C'est ce qui se passe dans l'émission induite : si N est très grand, une énorme augmentation d'intensité est possible, et, malgré sa faible probabilité dans le domaine optique, l'émission induite pourra néanmoins produire une amplification considérable du rayonnement.

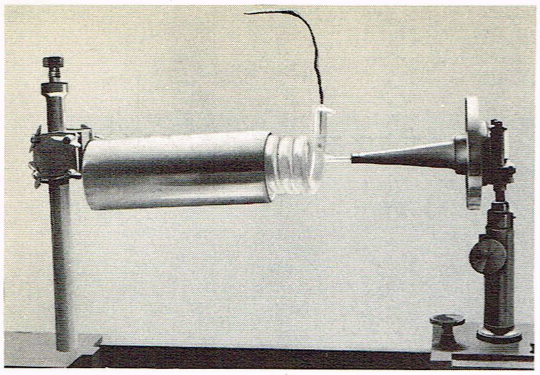
Illustration p. suivante.

Pour réaliser un laser, deux conditions sont finalement nécessaires : produire une inversion des populations pour que l'émission stimulée l'emporte sur l'absorption, et faire rayonner en phase un grand nombre d'atomes. L'un des matériaux les plus utilisés dans la fabrication des lasers est le rubis.

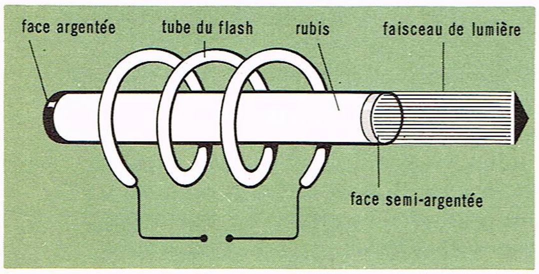
Les atomes sont représentés par des boules noires.



Cette figure montre comment varie la population des niveaux d'énergie E_1E_2. Les énergies sont portées en ordonnées et les populations correspondantes en abscisses. Plus le trait horizontal est long et plus le nombre N d'atomes se trouvant à ce niveau est important.

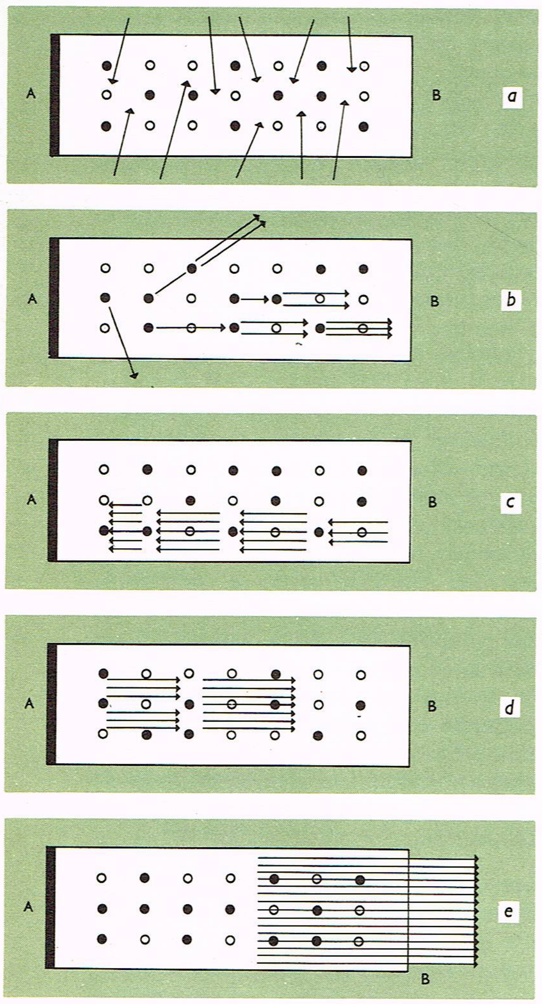


Photographie d'un laser à rubis et schéma.

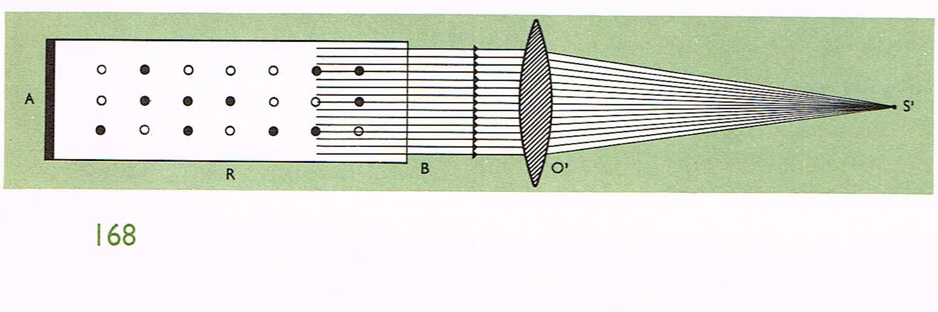


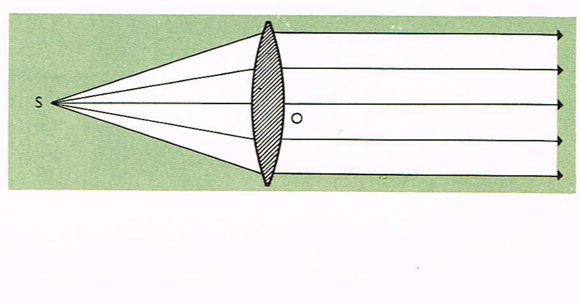
Pour inverser les populations, on utilise une méthode de pompage optique mise au point en France par A. Kastler. Le rubis est soumis à un rayonnement lumineux intense produit par un flash puissant. Sous son action, les atomes du rubis sont portés à l'état excité, et les conditions pour que l'émission induite l'emporte sur l'absorption sont remplies. Pour éviter les pertes par rayonnement, nous avons dit qu'il fallait placer le système dans une boîte dont les parois sont réfléchissantes. C'est le rubis luimême qui va constituer la boîte; on lui donne la forme d'une petit cylindre (de 3 mm de diamètre et 3 cm de longueur, par exemple), dont les extrémités, bien polies, sont planes et parallèles. L'une des extrémités est complètement argentée et l'autre semi-argentée. Le rubis ainsi préparé est entouré par le tube d'un flash qui va fournir l'énergie nécessaire au pompage. L'ensemble de ces dispositifs constitue un « laser à rubis ». Lorsque l'intensité de pompage du flash dépasse une certaine valeur, un faisceau de lumière rouge extrêmement intense sort de l'extrémité semi-argentée : le laser à rubis fonctionne.

Pour se donner une idée de la puissance d'un tel instrument, plaçons une lentille O′ à la sortie d'un laser et choisissons une focale telle qu'on obtienne en S′ une image de 0,1 mm de diamètre. Supposons que le flux est d'environ 4 millions de watts par centimètre carré (valeur réalisable). En focalisant le Soleil avec une lentille d'ouverture f/1 (diamètre égal à la focale, donc lentille de grande ouverture), on ne dépasse pas 500 watts par cm^2. **La petite image S′ produite par le laser au foyer de O′ est donc 10 000 fois plus brillante. Intensité lumineuse énorme, cohérence de phase et directivité sont les propriétés essentielles du laser.**



Les figures *a, b, c, d, e* permettent de comprendre le mécanisme de l'amplification par émission stimulée dans le rubis. Le rubis est fortement éclairé *(a)* et les atomes sont portés à l'état excité (boules noires : atomes à l'état fondamental; boules blanches : atomes à l'état excité). Quand un atome excité émet spontanément un photon, parallèlement à l'axe du rubis et représenté par une flèche *(b)*, le photon provoque l'émission induite par un autre atome. Le nouveau photon émis provoque à son tour l'émission induite par un autre atome, et le processus continue. Les photons émis dans d'autres directions sortent du rubis. L'amplitude de l'onde ainsi produite grandit continuellement. L'onde se réfléchit sur la face semi-argentée B et revient en arrière à travers la masse des atomes excités *(c)*. Son amplitude grandit toujours. Elle se réfléchit sur la face argentée A, revient sur B *(d)* et repart sur A, etc. Naturellement, il y a des pertes par réflexion sur B et même sur A, mais il arrive un moment où le gain produit par l'émission induite compense largement les pertes par réflexion sur les faces A et B. L'effet laser se produit : une onde sort par la face semi-argentée B *(e)*. La figure e montre que le faisceau qui sort du laser est presque un faisceau de rayons parallèles. Le laser se comporte un peu comme une source intense, mais très petite, S, placée au foyer d'une lentille O (*fig.* ci-dessous). Les rayons sortent presque parallèles; les rayons émergents divergent d'autant moins que pour une distance focale donnée SO de la lentille le diamètre de S est plus petit.



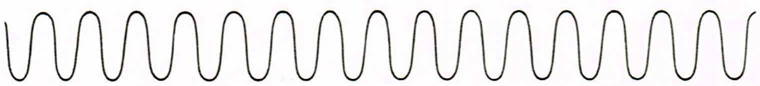


les multiples applications du laser

En concentrant à l'aide d'un objectif l'intensité considérable émise par un laser, on peut obtenir des effets thermiques intenses et localisés, par exemple produire localement des réactions chimiques ou effectuer des soudures de très petites dimensions. Une lame métallique, comme, par exemple, une lame de rasoir, peut être creusée et même percée. Les très fortes densités d'énergie lumineuse localisées ainsi conduisent à de nombreux effets inhabituels encore peu explorés. En particulier, des effets biologiques importants doivent pouvoir être obtenus de manière très localisée. Comme la focalisation d'énergie se fait sur un point infime, il devient possible de détruire une partie déterminée d'une seule cellule. L'extrapolation de ces conditions aux êtres vivants dans leur ensemble a fait parler du « rayon de la mort » émis par le laser; pour le moment, il ne s'agit heureusement que de tuer des microbes.

La directivité et la puissance du laser permettent de songer à en faire un véritable radar optique. La finesse du faisceau permet de choisir dans le paysage un objectif bien défini, un véhicule par exemple, et de mesurer instantanément et en plein jour sa distance.

Dans le domaine des communications, le laser peut opérer une véritable révolution. En effet, dans une source de lumière ordinaire, les atomes émettent des vibrations tout à fait au hasard (émission spontanée); on dit que les sources ordinaires émettent un rayonnement qui a le caractère d'un « bruit ». Ce défaut et la faiblesse de l'intensité des sources lumineuses ont rendu pratiquement impossible jusqu'ici l'emploi de la lumière dans les communications. Au contraire, avec un laser, on obtient une vibration presque parfaitement définie. La figure ci-dessous montre d'une manière très schématique les vibrations émises par une source ordinaire et par un laser. En modulant convenablement les vibrations émises par un laser, on peut transporter une énorme quantité d'informations, car il y a un nombre considérable de périodes de vibration lumineuse pendant une seconde.



Laser.

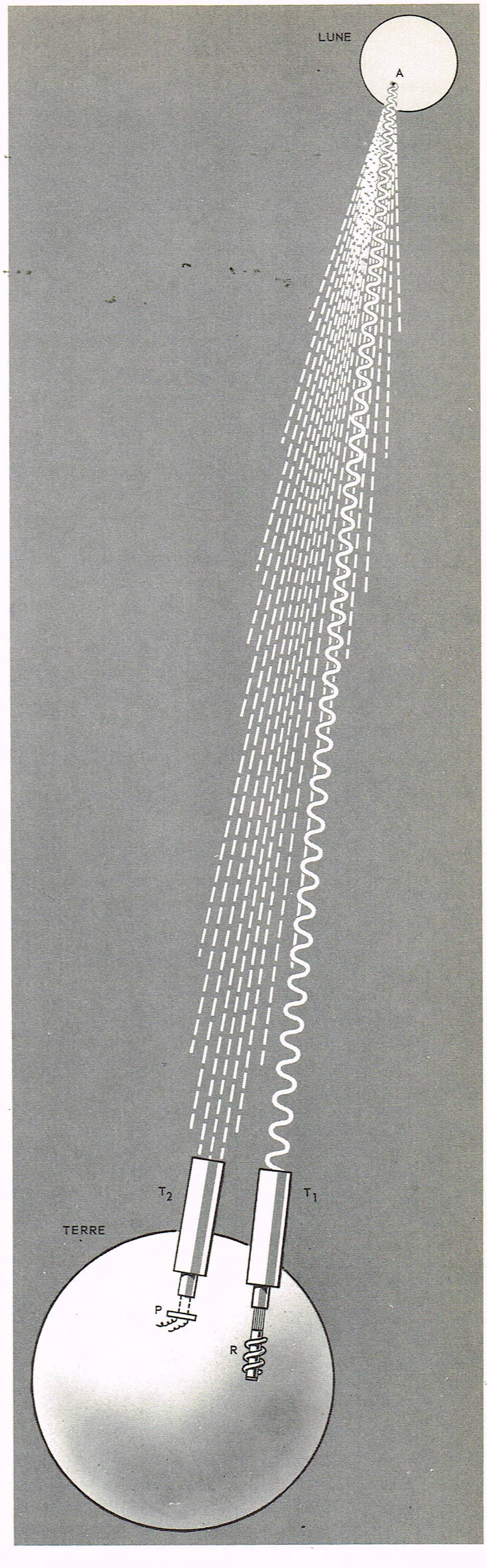


Source lumineuse ordinaire.

contact laser avec la Lune

La directivité et la puissance du laser permettent enfin d'envisager des transmissions spatiales. Pour ne pas perdre d'énergie, il y a intérêt à réduire encore la divergence, pourtant très faible, du faisceau émis par le laser. Pour cela, il suffit de placer à la sortie du laser un télescope fonctionnant en sens inverse du sens habituel. La lumière entre par l'oculaire et sort par l'objectif (le miroir). Le « contact » du faisceau avec la planète visée est observé avec un télescope muni d'un photomultiplicateur. L'expérience a été réussie pour la première fois en 1962, aux Etats-Unis. Le faisceau est dirigé sur la Lune par le télescope T_1 fonctionnant en sens inverse. Une petite région A de la Lune étant ainsi éclairée, la lumière diffusée est reçue par le télescope T_2 muni d'un récepteur sensible P (photomultiplicateur). Lorsque le laser fonctionne, le signal reçu par P prouve bien que la Lune a été éclairée depuis la Terre.

En perfectionnant les laser et les récepteurs, on peut envisager des portées utiles de l'ordre du milliard de kilomètres. **L'optique ouvre donc une ère nouvelle et prodigieuse dans le domaine des communications interplanétaires.**

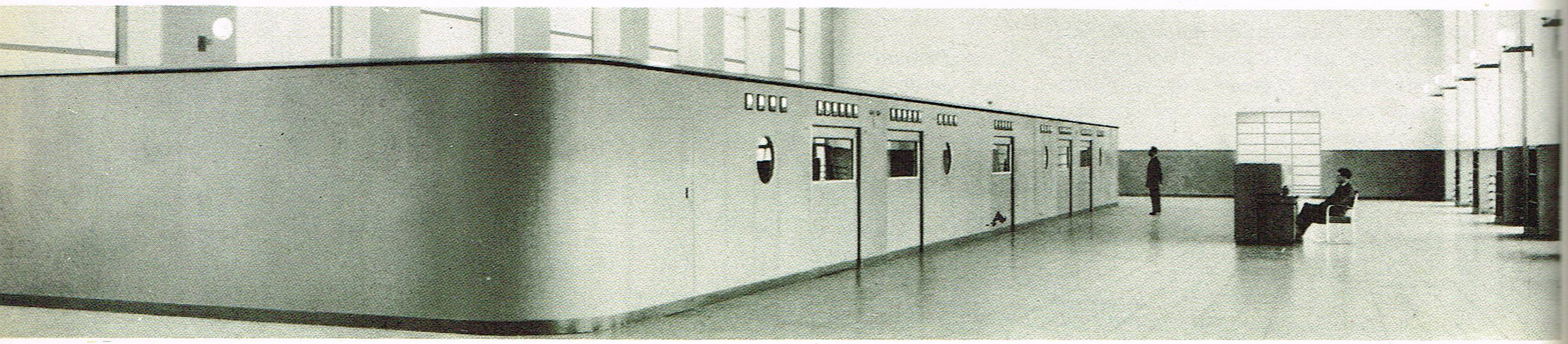




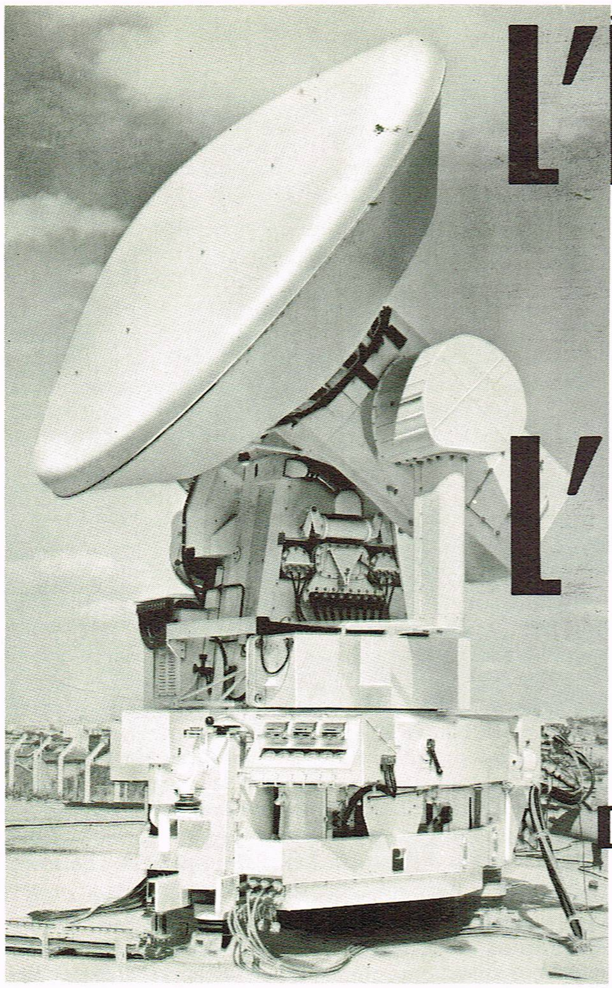
Les lignes simplifiées à l'extrême d'un poste émetteur de radiodiffusion moderne.



L'extraordinaire enchevêtrement d'antennes au sommet de la tour Eiffel.



Phot. C. S. F., René Bouillot et C. S. F., Georges Bru.



Radar traqueur de satellites artificiels.

L'ÉLECTRONIQUE outil de L'INFORMATION

par

Dominique LEPRINCE-RINGUET

Si vous voulez mettre dans l'embarras un de vos amis spécialiste de la question, essayez donc de lui faire définir l'électronique. Mais ne comptez pas trop sur ses réponses pour acquérir des idées claires : neuf fois sur dix, elles seront plus confuses encore qu'avant, et, vaguement inquiet, vous vous demanderez si c'est lui qui ne sait pas ce qu'il fait, ou vous qui êtes incapable de le comprendre.

Pour le non-initié, l'électronique, il faut bien le reconnaître, est particulièrement hermétique. Au prix d'un minimum d'effort, on peut comprendre le fonctionnement d'un moteur d'automobile, d'une fusée, d'un haut fourneau, parce que les opérations qui s'y déroulent parlent à notre imagination, à nos sens. Mais, en électronique, pas de support sensoriel : personne n'a jamais vu les électrons, et il n'est pas commode de se représenter leur mouvement.

Les difficultés que l'on éprouve à la définir ne facilitent pas la tâche de qui voudrait pénétrer dans le monde apparemment abstrait qu'elle constitue. Si l'on s'en tient à son nom, l'électronique devrait être la science, ou la technique, de l'électron. Mais elle n'est pas la seule qui s'en occupe : transformateurs, dynamos, moteurs, lampes d'éclairage ne relèvent pas de l'électronique bien qu'ils utilisent le courant électrique — c'est-à-dire un flux d'électrons —, mais de l'« électrotechnique ».

On a proposé de définir et de distinguer ces deux branches par l'objet qu'elles se proposent : fabriquer, transmettre et utiliser de l'énergie électrique, avec le meilleur rendement possible, pour l'électrotechnique; fabriquer, transmettre et recevoir de l' « information », en utilisant l'énergie électrique comme moyen, pour l'électronique. Mais qu'est-ce, exactement, que l'information? Est-il tellement plus aisé de définir cette notion récente que celle d'électronique ? Si bien qu'on n'est guère plus avancé (à cette question de l' « information » sera consacré un chapitre entier).

Essayons donc plutôt de cerner l'électronique en en dégageant les principaux aspects, les caractères spécifiques, tout ce qui contribue à la distinguer des autres techniques et des autres industries.

Phot. Thomson-Houston.

de quoi est fait un appareil électronique

Tout dispositif électronique consiste en un assemblage de pièces détachées élémentaires réunies entre elles de façon déterminée par du fil électrique. Les pièces détachées (ou « composants », pour leur donner leur appellation actuelle) sont groupées en familles, dont le nombre est relativement restreint. Les plus importantes sont constituées par les tubes à vide, les transistors, les diodes, les résistances, les capacités, les inductances.

L'ingénieur qui conçoit et réalise un appareil électronique aura donc à choisir les composants, puis à les interconnecter de façon convenable. Il agira un peu comme l'enfant qui construit une grue à partir de sa boîte de Meccano. Et, comme lui, si son montage ne lui donne pas satisfaction, il pourra sans difficulté le modifier : changer ou rajouter un composant est l'affaire d'un instant; il suffit de quelques coups de fer à souder.

Voici donc un caractère propre à l'électronique : l'expérimentation y est extrêmement aisée. Elle constitue même souvent, plus que projets et calculs, le moyen principal qui permet d'aboutir au matériel désiré.

En aéronautique, par exemple, l'ingénieur responsable de la fabrication d'un nouveau type d'avion devra attendre, si tout va bien, deux ou trois ans pour voir ses idées confirmées par l'expérience, et il sera impératif pour lui comme pour sa société (et pour le pilote d'essai!) que les idées en question soient valables.

Au contraire, l'électronicien qui se propose d'essayer un « circuit », c'est-à-dire un assemblage de composants, pourra vérifier presque séance tenante si celui-ci répond à ses prévisions; si ce n'est pas le cas, il n'en sera nullement troublé : il essaiera autre chose, n'ayant perdu qu'un minimum de temps. Avant d'aboutir à l'appareil définitif, poste de radio ou oscillographe cathodique, combien de « montages sur table », de maquettes, de prototypes ont fleuri sur la table d'expérimentation!

deux branches de l'électronique, deux activités bien différentes

De ce que nous avons dit plus haut, il ressort clairement qu'on peut, dans l'industrie électronique, distinguer deux branches principales : celle qui a trait à la fabrication des composants et celle qui a trait à la fabrication des dispositifs construits à partir des composants.

Les techniciens de la première sont d'ailleurs plus souvent des chimistes et des physiciens que des électroniciens proprement dits, ce qualificatif étant communément réservé à ceux de la seconde. Le fabricant de tubes à vide, s'il doit connaître les qualités que l'on attend de ses tubes, serait bien incapable de les utiliser dans des montages élaborés, de même que le « circuiteur » le plus expérimenté, tout en étant au courant des propriétés des tubes dont il se sert, aurait tout à apprendre s'il voulait aborder les problèmes délicats que pose leur fabrication.

De ces deux branches, c'est celle des composants qui est la plus importante : c'est elle qui sert de moteur à toute l'électronique, qui en conditionne les progrès. La mise au point de nouveaux composants actifs — nous verrons plus loin ce qu'il faut entendre par là — exige de véritables scientifiques (les inventeurs du transistor reçurent le prix Nobel), alors que la réalisation des appareils n'est qu'une technique, peut-être plus subtile, plus variée, plus changeante qu'une autre, mais qui ne peut donner lieu à des découvertes révolutionnaires. Il n'en était sans doute pas de même aux premiers temps de l'électronique, alors que la triode à vide existait déjà et qu'on n'imaginait pas encore tout ce qu'on pouvait en faire; mais, aujourd'hui, lorsqu'un nouveau composant apparaît, on sait presque tout de suite l'exploiter.

universalité de l'électronique

Ce qui fait peut-être la plus grande originalité de l'électronique, c'est son aptitude à trouver des applications dans toutes les industries, dans toutes les sciences. Un technicien de la métallurgie ne pourra guère trouver d'emploi en dehors de sa spécialité. **L'électronicien, lui, pourra s'engager dans l'industrie chimique, automobile ou aéronautique, dans un laboratoire de physique ou de biologie, voire dans un grand hôpital, et, évidemment, chez un fabricant de postes de télévision ou de magnétophones.**

L'électronique a envahi tous les domaines, et même les domaines extra-techniques. Les compagnies d'assurances, les banques ne font-elles pas appel aux machines à calculer arithmétiques? Les médecins n'utilisent-ils pas couramment l'électrocardiographe, et n'envisagent-ils pas de confier l'établissement des diagnostics à des machines? Avec la radio, la télévision, les électrophones, elle a pénétré dans les foyers, dont elle modifie les conditions de vie.

une évolution à rythme accéléré

Avant la Seconde Guerre mondiale, l'électronique s'intéressait presque exclusivement aux liaisons radio; on peut dire qu'elle est née en 1907, avec l'invention par l'Américain Lee De Forest de l' « Audion », première triode à vide, qui fut conçue dans le but d'améliorer la sensibilité de la détection des ondes hertziennes et resta confinée dans cet emploi pendant de longues années. Puis on apprit à faire de véritables amplificateurs, à utiliser la triode non seulement pour la réception, mais pour l'émission des ondes. Les premières liaisons avec des hyperfréquences apparurent, les premiers radars... Mais l'électronique restait indissolublement liée à la radio; on la désignait d'ailleurs plus volontiers par ce dernier terme.

Ce n'est vraiment que dans les années d'après guerre que son champ d'action s'est soudain étendu. Et avec quelle vitesse! Les télécommunications ont continué à se développer sur un rythme rapide, et, si leur dernier succès — les liaisons intercontinentales par l'intermédiaire d'un satellite — peut paraître extraordinaire, gageons qu'elles ne s'arrêteront pas là et que l'avenir, notamment grâce aux récents « masers » et « lasers », nous réservera bien des surprises. Parallèlement, la télévision est entrée dans les mœurs, et la télévision en couleurs s'apprête à prendre la relève. Les machines à calculer sont nées; les grandes calculatrices arithmétiques sont assurées de débouchés presque illimités et représentent déjà, dans les pays les plus évolués techniquement, une fraction très importante du marché.

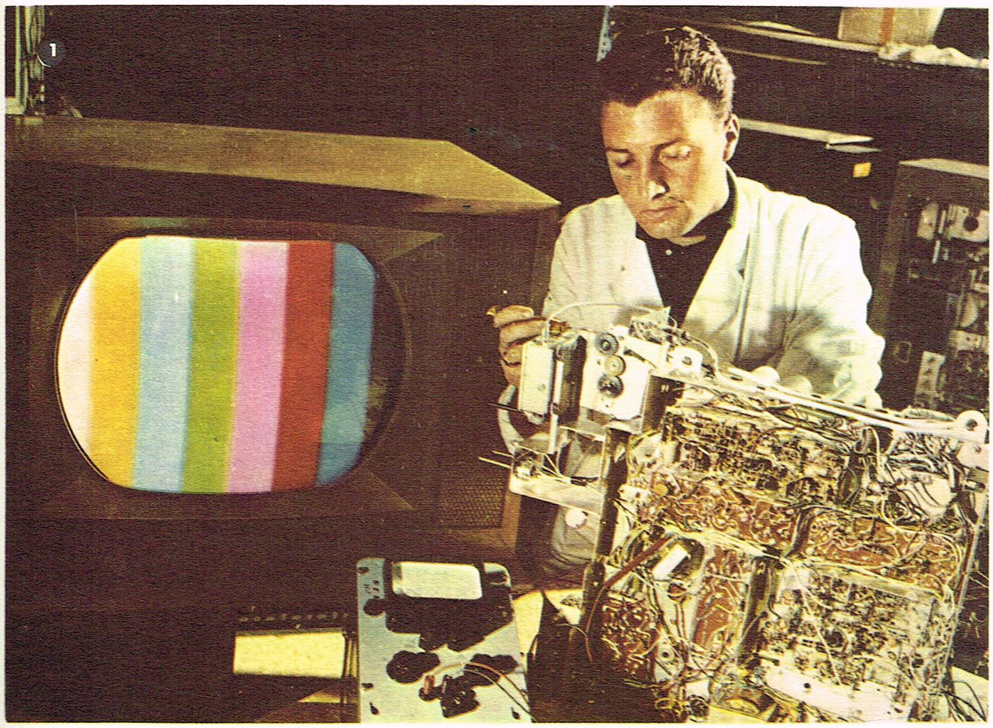
L'industrie a commencé à se tourner vers l'électronique, et non seulement par le biais des calculatrices : asservissement de trains de laminoirs, commande automatique de machines-outils, comptage d'objets, télévision industrielle. Là, l'évolution est moins rapide et n'en est sans doute qu'à ses débuts.

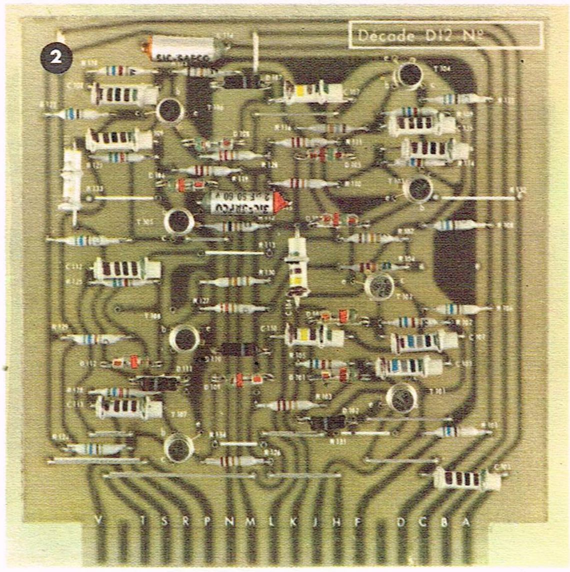
Enfin, comme on le sait, les armements militaires sont de plus en plus avides d'électronique : guidage d'avions et d'engins, calculateurs de tir, dispositifs de navigation (l'équipement électronique d'un chasseur peut représenter jusqu'à 50 p. 100 de son prix de fabrication; ce pourcentage ne fait que croître avec le temps et les perfectionnements des prototypes et engins).

de la galène aux semi-conducteurs

Si l'électronique a pu progresser de cette façon, si ses possibilités se sont constamment élargies, c'est grâce à la qualité toujours accrue de ses composants et à la mise au point de composants nouveaux.

L'avènement des semi-conducteurs, et en particulier du transistor à jonctions en 1952, a marqué un tournant; dépassant

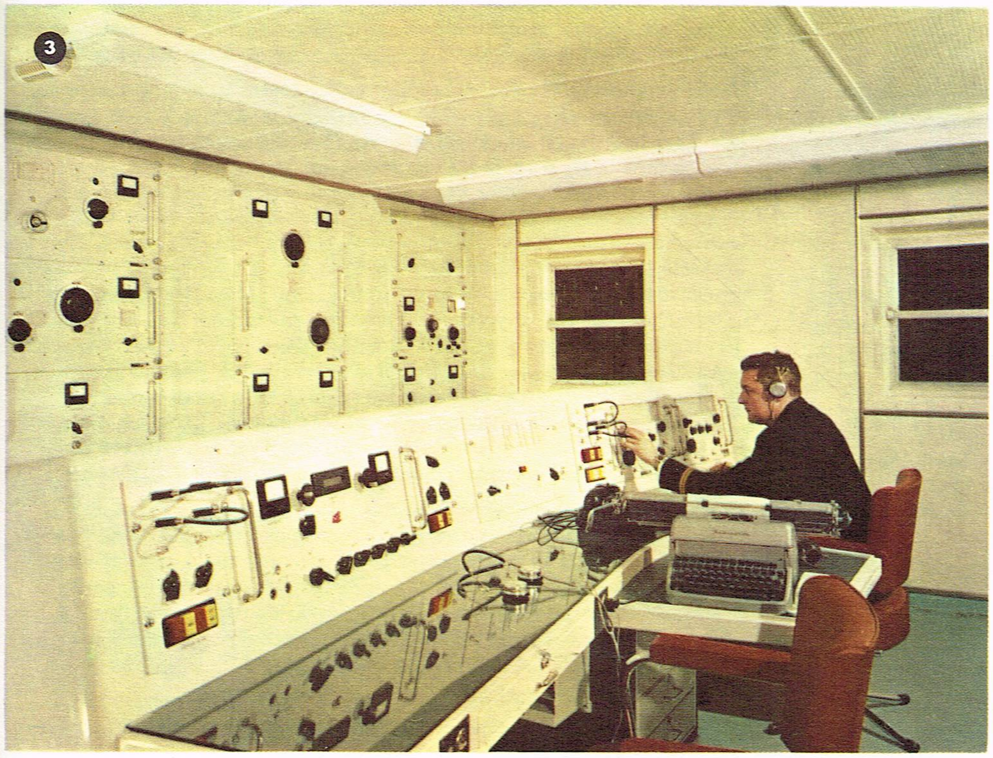




ÉLECTRONIQUE

1. **Réglage d'un téléviseur en couleurs procédé SECAM.**
2. **Montage sur circuit imprimé (on distingue le film de cuivre par transparence du support). [*Doc. Larousse.*]**
3. **Salle des télécommunications du paquebot « France ».** *(Phot. R. Bouillot C.S.F.)*
4. **Montage de tubes répétiteurs pour câbles sous-marins.** *(Phot. R. Bouillot C.S.F.)*
5. **Le transistor (désigné par le n° 137) donne l'échelle du microcircuit (en jaune) et des éléments qui le composent.** *(Phot. R. Bouillot C.S.F.)*







tous les pronostics, celui-ci a vu ses performances s'améliorer sans cesse et s'est affirmé comme le remplaçant du tube à vide classique, triode ou penthode. Beaucoup plus commode à utiliser, beaucoup plus sûr, il a ouvert quantité de portes à l'électronique; les machines arithmétiques, entre autres, lui sont redevables en grande partie de leur essor.

Mais de nombreux autres composants à semi-conducteurs ont vu le jour ces dix dernières années, tandis que les composants classiques se perfectionnaient encore.

Le rythme ne se ralentit nullement aujourd'hui, bien au contraire. Non seulement les caractéristiques, mais les prix, facteur non moins important, évoluent, souvent même à une allure vertigineuse; tel transistor « haute fréquence » qui fut mis sur le marché en 1960 pour 40 F, valait 15 F l'année d'après et 5 F en 1962.

Aussi l'électronicien doit-il constamment s'adapter, se maintenir au courant des nouveautés, remettre en question connaissances acquises et méthodes de travail, dans une discipline en si rapide évolution (tous les deux ans sort une machine électronique pour traiter l'information dix fois plus puissante que celles de la « génération précédente » : soit pratiquement un facteur 100 000 en dix ans).

Les sociétés elles-mêmes doivent faire preuve de beaucoup de souplesse et savoir modifier en temps utile leur orientation et leurs structures. Pour pouvoir suivre le mouvement, c'est-à-dire survivre, elles ont été amenées à gonfler leurs services d'études et de recherches. Le personnel en est particulièrement qualifié. Il est aussi, comme l'industrie électronique elle-même, particulièrement jeune.

composants passifs et composants actifs

Une première caractéristique des composants électroniques est leur extrême petitesse. Il existe certes des composants volumineux, utilisés en particulier lorsqu'on met en jeu des courants importants, mais généralement les courants sont faibles et les composants petits; comme nous le verrons, la tendance actuelle consiste à les rendre plus petits encore, à les « miniaturiser ».

Il est évident que tous ces éléments ne jouent pas, au sein d'un appareil, des rôles d'égale importance. C'est ainsi que l'on peut les classer, d'après leur fonction, en deux catégories : les éléments actifs et les éléments passifs. Les premiers (transistors, tubes à vide, etc.) sont spécifiquement électroniques, alors que les autres (résistances, inductances, condensateurs), s'ils sont indispensables à l'électronique, sont tout autant utilisés en électrotechnique.

On pourrait même, à partir de cette remarque, essayer de donner une nouvelle définition permettant de distinguer les deux sciences de l'électron : **il y a électronique s'il y a composant actif, sinon on a affaire à l'électrotechnique.**

Il s'agit maintenant de préciser la notion de composant actif, et pour cela parler de l'amplification.

une notion fondamentale : l'amplification

Tout le monde connaît la différence qui existe entre une automobile à direction classique et une automobile à direction « assistée ». Dans le premier cas, l'énergie nécessaire pour modifier la direction des roues avant est fournie entièrement par le pilote : lorsqu'il tourne le volant de la voiture, cette énergie est transmise jusqu'aux roues par une tringlerie et des pièces mécaniques, qui ne jouent qu'un rôle passif.

Dans le second cas, un élément moteur, plus précisément un servomoteur, est intercalé dans la transmission : le pilote n'a plus qu'une faible énergie à fournir, celle — beaucoup plus importante — qui est nécessaire au changement de direction des roues étant assurée par le servomoteur. Il y a eu amplification de l'énergie dépensée par le pilote; le servomoteur qui l'a réalisée a joué un rôle analogue à celui des composants actifs électroniques que nous cherchons à définir.

Mais, lorsqu'on parle d'amplification, il faut bien voir quelles sont les énergies mises en jeu. Il y a d'abord l'énergie de commande, ou énergie d'entrée (ici, celle qui est fournie par le pilote), puis l'énergie d'utilisation, ou énergie de sortie (celle qui permet de modifier la direction des roues). Enfin, il existe évidemment une troisième énergie — l'énergie auxiliaire — qu'il faut fournir au système amplificateur, car celui-ci ne saurait fournir plus d'énergie qu'il n'en reçoit. Comme il en consomme lui-même toujours un peu, l'énergie auxiliaire sera même supérieure à la différence des deux autres.

Dans le cas de la direction assistée de la DS 19, dotée d'un servomoteur hydraulique, l'énergie auxiliaire est fournie par des pompes entraînées par le moteur de la voiture; c'est ce dernier qui, tout compte fait, effectue le travail commandé par le pilote.

En somme, l'amplification ne consiste pas à fabriquer une grande énergie à partir d'une petite, mais plutôt à diriger une fraction d'une énergie disponible vers l'utilisation, éventuellement à contrôler l'importance de cette fraction, et ceci en dépensant une énergie plus faible.

les composants actifs permettent l'amplification de l'énergie électrique

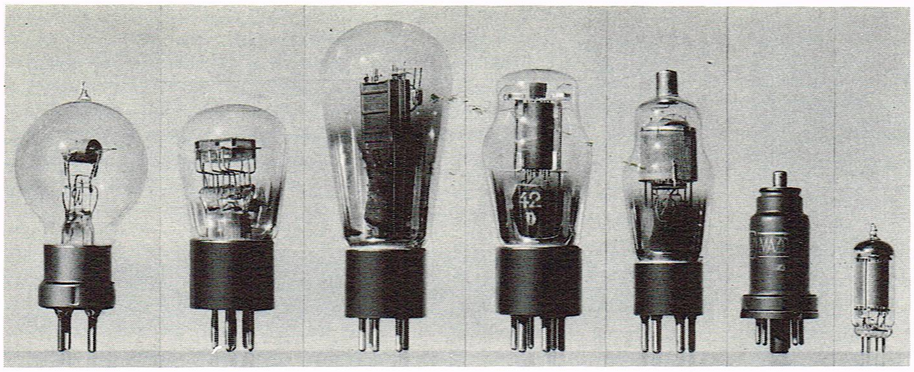
Revenons à l'électronique. On dit qu'un composant est « actif » lorsqu'il permet d'amplifier une énergie électrique. Et ce directement, sans passer par l'intermédiaire d'une autre forme d'énergie. L'énergie auxiliaire est alors fournie par une ou plusieurs sources de tension : piles, accumulateurs ou secteur.

Dans l'exemple illustré ci-contre, on constate que le « rendement » de l'amplificateur n'est pas fameux, puisqu'il faut dépenser 22,5 mW pour n'en utiliser que 12,5. Il eût été encore bien plus mauvais si, à la place du transistor, on avait employé un tube à vide. Mais la notion de rendement n'est pas le souci majeur des électroniciens; c'est un problème qui peut avoir son importance, mais qui reste secondaire, alors qu'en électrotechnique c'est le problème numéro un. On se préoccupera plutôt de la qualité de l'amplification : pour un amplificateur proportionnel, on cherchera à rendre le facteur de proportionnalité aussi constant que possible dans des limites aussi étendues que possible; si le signal d'entrée varie dans le temps, on essaiera de conserver l'amplification même si les variations sont très rapides, si la fréquence est très élevée. On peut d'ailleurs remarquer que **plus la qualité d'un amplificateur est bonne, plus son rendement énergétique est mauvais.**

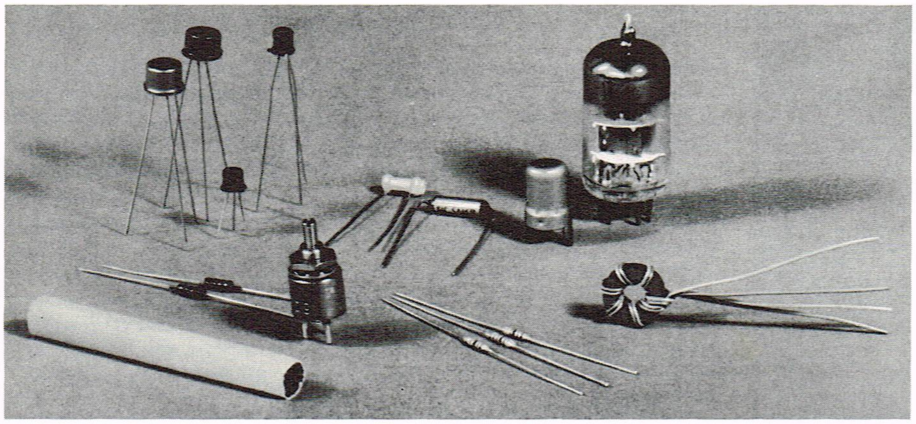
Les composants actifs sont donc ceux qui permettent de réaliser une amplification de puissance électrique. Et c'est bien de puissance qu'il s'agit. Un transformateur permet certes d'amplifier une tension, mais non une énergie : la puissance de sortie n'est jamais supérieure à celle d'entrée (la différence sert à chauffer inutilement et fâcheusement l'appareil, qu'il faut souvent refroidir); c'est un composant passif.

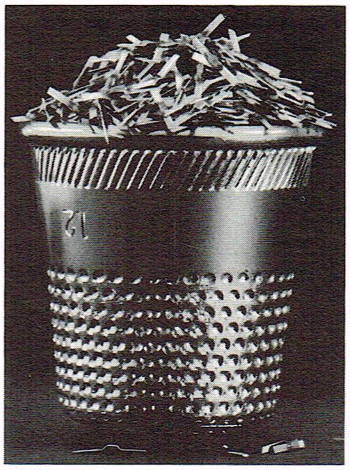
Ceci ne veut pas dire que tous les composants actifs sont montés en amplificateurs. Ils peuvent être utilisés par exemple comme

L'évolution des techniques a conduit notamment à une rapide miniaturisation des composants.



Ci-contre, **évolution des tubes à vide depuis le tube TM (télégraphie militaire) de 1916.** *En dessous*, **le tube miniature fait déjà figure de géant à côté des diodes, transistors, condensateurs et résistances actuels. (La cigarette donne l'échelle.)**





Ce dé ordinaire contient 1 053 diodes remplaçant autant de grosses lampes.

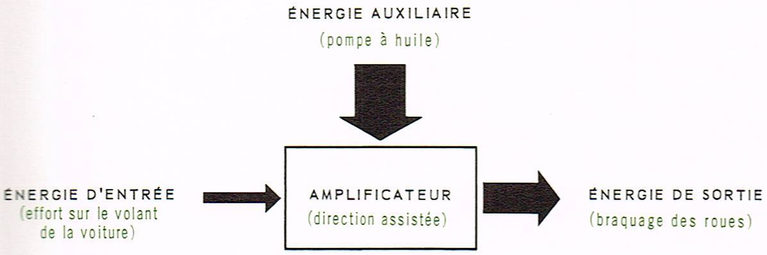
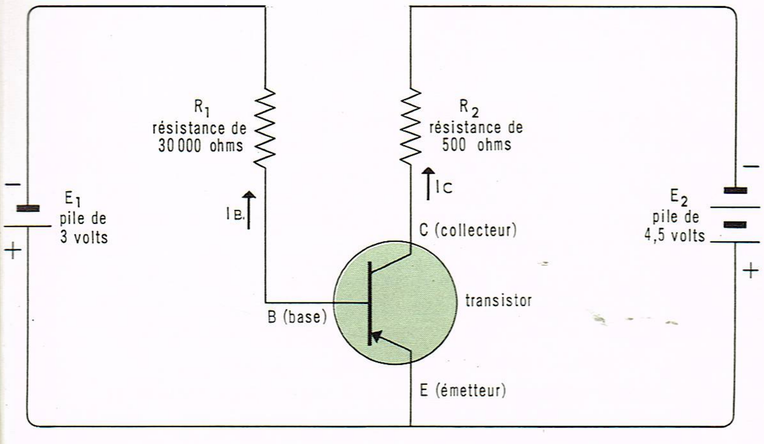


Schéma de l'amplification et, en vert, analogie de l'amplificateur avec un système de direction assistée.

Le système amplificateur consomme lui-même de l'énergie, ce qui fait que l'énergie de sortie est inférieure à la somme de l'énergie auxiliaire et de l'énergie d'entrée.



TRANSISTOR MONTÉ EN AMPLIFICATEUR

Il existe un rapport constant entre les courants I_B, qui circule dans la base, et I_C, qui passe dans le collecteur. Ce rapport, appelé « gain en courant du transistor », varie, suivant les types, entre quelques dizaines et quelques centaines. Supposons donc que notre transistor ait un gain de 50, ce qui veut dire que le courant de collecteur est 50 fois plus grand que le courant de base.
La puissance d'entrée (à amplifier) est ici obtenue à partir du générateur de tension E_1 (pile de 3 V), la puissance auxiliaire à partir du générateur E_2 (pile de 4,5 V), la puissance de sortie, ou puissance d'utilisation, est celle qui est dissipée dans la résistance R_2, placée dans le collecteur du transistor. Ce pourra être, par exemple, la bobine d'un relais électromécanique qu'on cherchera à commander avec une puissance plus faible que celle qui serait nécessaire si on l'attaquait directement.
En négligeant la différence de tension entre base et émetteur, toujours très faible, on peut calculer le courant de base : $I_B = \frac{E_1}{R_1} = \frac{3}{30\,000}$ A $= 10^{-4}$ A ou 0,1 mA; et la puissance d'entrée : $P_1 = E_1 \times I_B = 3 \cdot 10^{-4}$ W ou 0,3 mW.
Le gain en courant du transistor étant ici de 50, le courant de collecteur sera $I_C = 50 \times I_B = 5 \cdot 10^{-3}$ A, et la puissance qui apparaît dans la résistance R_2 (puissance de sortie) est $P_2 = R_2 I_C^2 = 500 \times 25 \cdot 10^{-6}$ W ou 12,5 mW.
Quant à la puissance auxiliaire, elle est $P_3 = E_2 \times I_c = 4{,}5 \times 5 \cdot 10^{-3}$ W $= 22{,}5$ mW.

Phot. Boiron (Mazda), C. R. C., T. R. W.

générateurs d'oscillations entretenues (oscillateurs) ou dans les « circuits logiques » des machines à calculer. Ce que l'on peut affirmer, c'est que, dans tous les cas, ils ont à réaliser une amplification de puissance pour remplir leur fonction.

ce que sont les composants actifs

Les composants actifs comprennent en gros deux grandes catégories : les composants à tubes et ceux à semi-conducteurs. Dans les premiers, les électrons se déplacent dans le vide ou dans un gaz, alors que dans les autres ils se déplacent dans un corps cristallin solide. Les tubes ont régné sur l'électronique depuis sa naissance (invention de la triode par Lee De Forest en 1907) jusqu'à ces dernières années. Si, maintenant encore, leur place y est importante, ils sont néanmoins dépassés, et le seront de plus en plus, par les semi-conducteurs, dont l'existence en tant que composants actifs ne remonte qu'à 1948.

Ils comprennent les tubes à vide (triodes, tétrodes, penthodes) d'emploi très universel et des tubes à usages plus particuliers : thyratrons (tubes à vapeur de mercure), tubes « hyperfréquence » (magnétrons, klystrons, tubes à ondes progressives) utilisés pour le radar et les télécommunications. Les composants actifs à semi-conducteurs sont encore trop récents pour que leurs différentes « races » soient bien stabilisées. Des dispositifs nouveaux apparaissent très fréquemment; certains semblent promis à un grand avenir, et pourtant disparaissent par la suite, remplacés par d'autres plus intéressants ou plus faciles à fabriquer. **Seul le transistor (triode à jonctions) a désormais acquis une structure qui semble définitive; comme les triodes et penthodes à vide qu'il tend à remplacer, c'est le composant à tout faire de l'électronique classique.** Ses différents modèles permettent son utilisation à faible ou forte puissance, aux très faibles courants, aux fréquences élevées, et lui confèrent un caractère universel.

Parmi les autres composants à semi-conducteurs, citons les redresseurs contrôlés, de fonction équivalente à celle des thyratrons à gaz, les diodes trijonctions et les transistors unijonctions, les dispositifs à « effet de champ », dont les caractéristiques se rapprochent de celles des penthodes à vide, et surtout les diodes « tunnel », très intéressantes aux fréquences élevées.

Enfin, tout récemment sont apparus des dispositifs fondés sur des principes différents et qui permettent l'émission et l'amplification d'ondes électromagnétiques de fréquences beaucoup plus élevées que celles auxquelles on était arrivé jusqu'à présent, et même d'ondes lumineuses. **Il s'agit des masers et des lasers, déjà très employés industriellement, qui semblent devoir permettre à l'électronique de faire un nouveau pas en avant.**

les semi-conducteurs

Nous commencerons l'étude des composants électroniques par les dispositifs à semi-conducteurs, et en particulier par le plus important de ceux-ci, le transistor à jonctions.

On peut se faire une idée de l'importance relative prise par les composants à semi-conducteurs par les constatations suivantes : en 1950, le chiffre d'affaires qu'ils représentaient aux Etats-Unis était négligeable devant celui qui était réalisé sur les composants à tubes; en 1961, on a pu calculer que ce chiffre d'affaires était, dans le même pays, de 606 millions de dollars, contre 593 pour l'ensemble des composants à tubes. Les deux grandes familles étaient donc alors à peu près à égalité; mais la première continue à se développer à un rythme accéléré (il y a actuellement aux Etats-Unis une centaine de compagnies qui fabriquent des semi-conducteurs), alors que la deuxième est pratiquement stationnaire; de plus, tout ce qui est nouveau en électronique classique se fait avec des semi-conducteurs, et, dans le chiffre d'affaires relatif aux tubes, entre pour une bonne part ce qui sert à la maintenance de matériels déjà relativement anciens appelés à disparaître progressivement. **On peut donc affirmer d'ores et déjà que les semi-conducteurs ont supplanté les tubes, et c'est pourquoi nous les faisons passer ici au premier rang.**

de tout petits éléments, et presque rien dedans

Nous avons vu quelques échantillons de transistors et pu remarquer les faibles dimensions que pouvait avoir ce composant, pourtant si important. Mais si nous ouvrons le boîtier d'un transistor, notre étonnement fait place à la stupéfaction : il est presque complètement vide! La partie utile du transistor, celle où circule le courant, est constituée par le tout petit volume de matière où sont connectés les trois fils de commande, le boîtier n'étant là que pour entourer le transistor d'une enveloppe gazeuse qui le protège des impuretés atmosphériques et pour permettre, en jouant le rôle de radiateur, une meilleure évacuation de la chaleur qui apparaît lors du passage du courant. Tous les transistors ne sont pas aussi petits. Mais on est toujours frappé par les faibles dimensions de la partie utile, même lorsqu'il s'agit de transistors de puissance.

Cette partie utile est faite de semi-conducteurs. Essayons donc d'abord de définir les semi-conducteurs et de décrire leurs principales propriétés. Nous verrons ainsi : qu'on peut faire, à partir du germanium ou du silicium, deux sortes de semi-conducteurs : le type « N » et le type « P »; qu'un élément N, convenablement juxtaposé avec son homologue P, forme une « jonction » redresseuse, c'est-à-dire qui ne laisse passer le courant que dans un sens; enfin qu'un transistor est constitué par un fragment d'un type compris entre deux fragments de l'autre type; il y a donc deux jonctions dans un transistor.

la semi-conduction, chapitre récent de la physique du solide

Nous avons tous appris, dans les manuels de physique, qu'on peut classer les corps en deux catégories, du point de vue de leurs propriétés électriques : les conducteurs d'une part, les isolants d'autre part. En revanche, on ne nous a guère parlé des semi-conducteurs, catégorie intermédiaire, parce qu'ils ne possédaient pas encore un grand intérêt pratique, et qu'on ne s'expliquait pas clairement leurs propriétés.

C'est à la physique du solide qu'il faut demander l'explication du phénomène de semi-conduction, qui s'appuie notamment sur la notion de « bandes d'énergie » dans un cristal. Nous n'en exposerons pas ici la théorie, qui sera traitée très en détail au tome II, et nous nous contenterons de faire appel à des notions plus simples et plus intuitives.

Les semi-conducteurs qui nous intéressent sont constitués par un cristal aussi pur que possible, dans lequel on a introduit une quantité très faible, mais néanmoins parfaitement dosée, d'un corps étranger appelé « impureté ».

Le cristal est constitué le plus souvent par un corps simple occupant la quatrième colonne dans la classification pério-

dique des éléments, c'est-à-dire un corps de valence 4, par exemple le germanium ou le silicium.

Les impuretés sont de deux types. Ou bien ce sont des corps simples occupant la troisième colonne du même tableau (valence 3) — on a affaire à un semi-conducteur de type P —; ou bien ce sont des corps simples de la cinquième colonne (valence 5) — il s'agit alors d'un semi-conducteur de type N.

disposition des atomes dans un cristal de germanium ou de silicium

Pour simplifier le texte, nous parlerons dans ce qui va suivre uniquement du germanium, les raisonnements qu'on pourrait faire à propos du silicium étant identiques.

Un atome, rappelons-le, est constitué par un noyau très dense chargé positivement, autour duquel gravitent un certain nombre d'électrons, dont la charge globale — négative — est égale à la charge du noyau. Ces électrons ont des niveaux d'énergie bien précis : on peut dire qu'ils sont répartis sur des « couches » situées à des distances du noyau bien définies et en nombre restreint, chaque couche correspondant à une énergie donnée. En outre, il ne peut y avoir plus d'un certain nombre d'électrons sur une même couche; les électrons d'un atome « remplissent » d'abord les couches les plus proches du noyau, la couche extérieure étant en général partiellement remplie. Dans le cas du germanium, il y a quatre électrons sur la couche externe, celle-ci étant complète avec huit électrons.

Lorsque ce corps se présente sous la forme d'un cristal pur, ses atomes sont répartis dans l'espace de façon régulière, à égale distance les uns des autres. Chacun est réuni aux quatre atomes les plus proches par des forces de liaison : on peut dire qu'il est placé au centre d'un tétraèdre dont les sommets sont occupés par ces atomes.

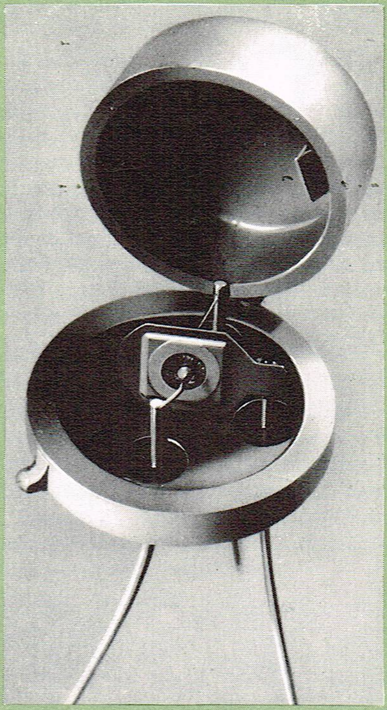
Les liaisons d'un atome avec un atome voisin consistent en particulier en la mise en commun de deux électrons de la couche périphérique : un atome « prêtera » un électron à son voisin, et en même temps lui « empruntera » un autre électron; de ces deux électrons, il ne sera d'ailleurs pas possible de dire lequel appartient au premier atome, et lequel à l'autre. Comme chaque atome est lié à quatre autres atomes, il sera finalement entouré de huit électrons, dont il aura fourni la moitié.

parfaitement pur, le cristal est un isolant

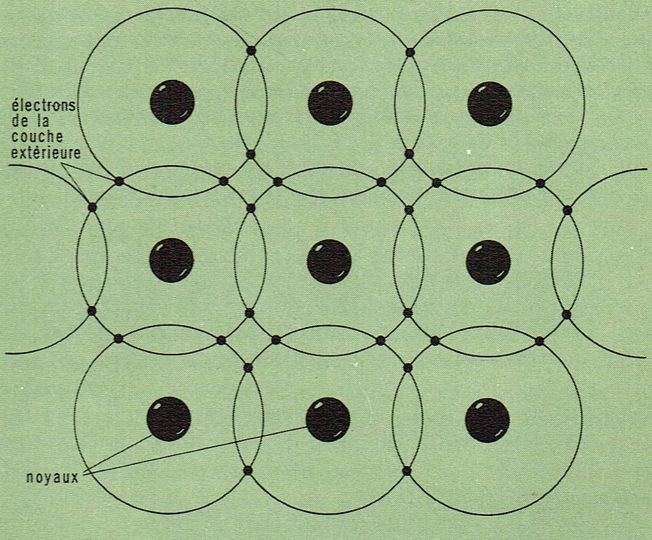
Ainsi, chaque atome du cristal a complété sa couche externe à l'aide des atomes voisins, ce qui fait que toutes les couches sont complètes, puisque, nous l'avons vu, les couches internes le sont de toutes façons (au lieu de « couches », il faudrait en fait parler ici de « bandes d'énergie », les niveaux d'énergie des électrons n'étant pas les mêmes dans le cristal et autour de l'atome isolé).

Or, on montre que les électrons qui appartiennent à une couche pleine ne peuvent se déplacer dans le cristal sous l'action d'un champ électrique; autrement dit, ils ne peuvent lui conférer un caractère conducteur. Dans notre cristal, comme toutes les couches sont pleines, aucun électron ne pourra se déplacer; nous avons donc affaire à un isolant.

Cependant, deux phénomènes sont susceptibles de modifier cette nature isolante du cristal de germanium, et celui-ci pourra, dans une certaine mesure, se transformer en un conducteur de l'électricité. Ces deux phénomènes sont : l'action de la température d'une part, qui donnera lieu à une conductivité « intrinsèque », et l'adjonction de certaines impuretés d'autre part, qui provoquera une conductivité « extrinsèque ».



Intérieur du transistor.
Le transistor proprement dit est constitué sur le petit rectangle vertical, au centre du boîtier. L'arrivée des trois fils d'émetteur, de base et de collecteur est bien visible, ainsi que leur connexion au transistor.



Représentation plane d'un cristal de germanium ou de silicium. Chaque atome, normalement entouré de 4 électrons périphériques, met ceux-ci en commun avec les 4 atomes voisins. La couche externe, ainsi pourvue de 8 électrons, est alors « complète » et fait d'un cristal de ce type un isolant.
(Les électrons de la couche externe ont été représentés sur un cercle ayant pour centre le noyau; ceux des couches internes n'ont pas été figurés.)

chauffons : il apparaît des électrons libres et des trous positifs

Tout ce que nous avons dit jusqu'à présent sur la structure du cristal de germanium n'est vrai, en toute rigueur, qu'au zéro absolu. Sous l'action de la température, il apparaît des déformations dans le réseau cristallin, qui se traduisent par des vibrations des atomes, lesquelles communiquent aux électrons de l'énergie supplémentaire. Ceux-ci, bien que fortement liés aux atomes, vont alors de temps en temps rompre leur liaison et s'échapper dans le réseau, où ils pourront se déplacer librement et participer à la conduction. Plus élevée sera la température, plus il y aura d'électrons libres, et plus conducteur sera le germanium.

Les électrons de certains autres corps sont davantage liés aux atomes, et l'action de la température ne leur communique pas une énergie suffisante pour rompre leurs liaisons; ces corps sont des isolants.

Quant aux conducteurs, ce sont des corps dont les atomes, disposés en réseau cristallin, ne complètent pas leur couche périphérique. Les électrons appartenant à cette couche sont libres de se déplacer dans le réseau, quelle que soit la température.

Revenons au cas du germanium. Lorsqu'un électron, sous l'effet de la température, a été arraché aux atomes auxquels il était lié, il a laissé derrière lui une place vacante, à laquelle est associée une charge positive, égale en valeur absolue à la charge de l'électron (puisque l'ensemble noyau + électrons d'un atome est électriquement neutre). Cette place vacante, encore appelée « lacune » ou « trou positif », pourra facilement être occupée de nouveau par un électron d'un atome voisin, lequel laissera un nouveau « trou » à l'endroit qu'il a quitté. Les trous peuvent donc se déplacer dans le réseau de proche en proche, et la théorie montre que leurs mouvements sont régis par les mêmes lois que ceux des électrons libérés par l'agitation thermique.

Bien entendu, les électrons libérés peuvent de temps en temps réoccuper des places vacantes et retomber ainsi sous l'action des forces qui les lient aux atomes; on dit qu'il y a « recombinaison » entre électrons libres et trous. A une température donnée, un équilibre s'établit, et le nombre d'électrons libérés — donc de trous — est constant; cela veut dire que chaque fois qu'un nouvel électron sera arraché au réseau son apparition sera, en moyenne, compensée par une recombinaison.

En résumé, la température a pour effet de faire apparaître, dans un cristal de germanium parfaitement pur, deux types de particules susceptibles de s'y déplacer librement : des électrons chargés négativement, et des trous, en nombre égal, chargés positivement. On dit qu'il y a création de « paires » électron-trou.

A la température ordinaire, le germanium est encore très peu conducteur; sa conductivité à 25 °C (298 °K) est environ 10 millions de fois plus faible que celle du cuivre, mais elle augmente exponentiellement avec la température et commence à devenir sensible vers 80 °C (353 °K).

on peut aussi rendre un cristal conducteur en lui ajoutant un nombre infime d'impuretés pentavalentes

Examinons ce qui se passe lorsque sont présents dans le cristal un petit nombre d'atomes d'un corps simple (arsenic, phosphore ou antimoine) possédant cinq électrons sur sa couche périphérique. La proportion d'atomes d'impureté par rapport aux atomes de germanium étant très faible, la forme du réseau cristallin ne sera pas modifiée, chaque atome d'impureté venant prendre de temps à autre la place d'un atome de germanium. Comme lui, il échangera avec chacun de ses quatre voisins — qui seront tous des atomes de germanium — un électron périphérique, ce qui fait qu'il sera entouré de la couche à huit électrons caractéristique des atomes du cristal, avec toutefois un électron supplémentaire, puisqu'il a fourni seulement quatre électrons à ses voisins et qu'il en possédait cinq à l'origine.

Or, s'il se trouve que la couche complète à huit électrons est très stable et fortement liée aux atomes, l'électron supplémentaire l'est beaucoup moins. Ce qui veut dire que l'énergie nécessaire pour l'arracher à l'atome d'impureté auquel il appartient est beaucoup plus faible que celle qu'il faut fournir pour arracher un électron à un atome de germanium. Pour fixer les idées, disons qu'elle est près de cent fois plus petite.

Au zéro absolu, tous les électrons resteront liés aux atomes, et le germanium, malgré la présence des impuretés, sera parfaitement isolant; mais, dès que la température augmentera un peu, l'énergie communiquée aux électrons sera suffisante pour libérer le cinquième électron des atomes d'impureté.

Celui-ci se déplacera alors sans contrainte dans le réseau, comme, tout à l'heure, les électrons du germanium. Mais, alors que les électrons appartenant aux atomes de germanium laissent derrière eux, lorsqu'ils sont libérés par l'effet de la température, des « trous » aussi mobiles qu'eux, les électrons des atomes d'impureté abandonnent, une fois arrachés, des ions positifs fixes; en effet, l'atome d'impureté qui a perdu un électron acquiert une charge positive égale en valeur absolue à celle de l'électron : c'est un ion. D'autre part, il n'a pratiquement aucune chance de récupérer un électron qui lui ferait perdre sa charge, puisque, dès que la température est assez élevée, tous les électrons supplémentaires sont libérés. C'est notamment le cas à la température ordinaire; il n'y a donc pas l'équivalent des « recombinaisons » que nous avons observées tout à l'heure.

Ainsi, l'adjonction d'impuretés à cinq électrons de valence a pour effet de faire apparaître dans le réseau, dès que la température est un peu supérieure au zéro absolu, d'une part des électrons libres, susceptibles de rendre le germanium conducteur, d'autre part des ions positifs fixes. A la température ordinaire, tous les atomes d'impureté sont ionisés, ce qui veut dire qu'il y a autant d'électrons libres que d'atomes d'impureté. La conductivité du germanium qui en résulte, ou conductivité extrinsèque, étant proportionnelle au nombre d'électrons libres, est bien déterminée par le nombre d'atomes d'impureté. Même lorsque la proportion de ces derniers est faible (on prend souvent des taux de l'ordre de 1 pour 1 million), la conductivité extrinsèque est bien plus grande — à la température ordinaire — que la conductivité intrinsèque, c'est-à-dire celle qui est due à l'action de la température sur les seuls atomes de germanium.

Nous verrons que c'est la conductivité extrinsèque qui permet de réaliser les diodes et les transistors, la conductivité intrinsèque étant au contraire un phénomène parasite. Or, cette dernière croît exponentiellement avec la température alors que la première reste constante aux températures usuelles. On peut donc, dès maintenant, prévoir que les dispositifs à semi-conducteurs ne peuvent pas fonctionner à trop haute température.

les impuretés trivalentes sont également efficaces

Supposons maintenant que notre cristal contienne un petit nombre d'atomes, non plus à cinq électrons périphériques, mais à trois, tels que l'indium, l'aluminium, le bore.

Comme tout à l'heure, chaque atome de ce nouveau type d'impureté viendra prendre la place d'un atome de germanium dans le réseau cristallin sans en modifier la structure géométrique; il mettra en commun avec ses quatre voisins les électrons de sa couche externe. Mais, comme il n'a que trois électrons à offrir, il devrait normalement être entouré par une couche

à sept électrons. C'est effectivement ce qui se passe au zéro absolu : le germanium à impuretés trivalentes est alors parfaitement isolant, tout comme le germanium à impuretés pentavalentes.

Cependant, cette couche à sept électrons aura fortement tendance à se compléter à huit, et ce sera chose faite dès que la température sera un peu supérieure au zéro absolu. Mais où donc l'atome d'impureté prendra-t-il ce huitième électron, va-t-on se demander. Eh bien, tout simplement à un atome de germanium voisin. En effet, s'il faut communiquer beaucoup d'énergie aux électrons du germanium pour qu'ils s'échappent dans le réseau, on montre qu'il en faut beaucoup moins (près de cent fois moins) pour que l'un d'eux vienne compléter à huit la couche à sept électrons dont est entouré l'atome d'impureté. En quittant son atome de germanium, l'électron en question laissera derrière lui une place vacante, qui sera bientôt elle-même occupée par un électron appartenant à un autre atome de germanium, et ainsi de suite. Cette place vacante constitue un trou positif, et nous avons vu tout à l'heure qu'elle était libre de se déplacer dans le réseau et de participer à la conduction, exactement de la même façon que les électrons arrachés aux atomes d'impureté à cinq électrons de valence.

Quant à l'atome d'impureté trivalente, il aura toutes les chances de garder l'électron supplémentaire qu'il a récupéré, et constituera un ion négatif fixe.

Résumons-nous : lorsque le cristal de germanium est « dopé » par des impuretés à cinq électrons de valence, il apparaît, dès que la température est suffisante, des électrons libres en nombre égal à celui des atomes d'impureté. Ceux-ci deviennent alors des ions positifs fixes; le germanium est dit « de type N » parce que sa conductivité est assurée par des électrons de charge négative.

Lorsque les impuretés sont des corps à trois électrons de valence, il apparaît, dans les mêmes conditions, des charges positives mobiles, ou trous, en nombre égal à celui des atomes d'impureté, qui deviennent des ions négatifs fixes. Le germanium est dit « de type P » parce que sa conductivité est assurée par des charges positives.

Les trous jouissent des mêmes propriétés que les électrons libres; en particulier leur mobilité est du même ordre de grandeur.

deux formes de courant

La première méthode pour faire naître un courant dans un barreau de germanium N ou P consiste à le placer dans un champ électrique.

On sait, en effet, qu'une charge électrique placée dans un champ est soumise à une force proportionnelle à la valeur de la charge et à l'intensité du champ. Si elle peut se déplacer, elle sera animée d'un certain mouvement.

Dans le cas du germanium N (ou P), les électrons libres (ou les trous) vont donc se déplacer sous l'action d'un champ. On pourra mettre ce phénomène en évidence en prenant un barreau de germanium dopé et en reliant ses extrémités aux électrodes d'une pile électrique; les charges libres, en s'y déplaçant, donneront naissance à un courant.

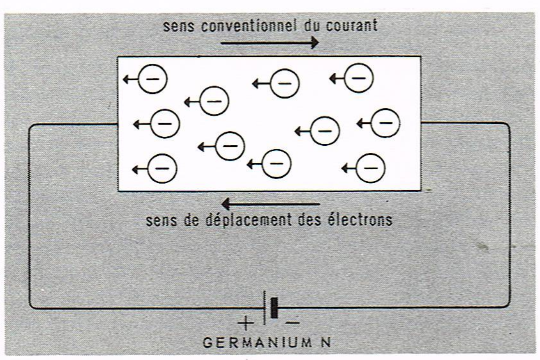
Remarquons ici qu'un électron soumis à l'action d'un champ électrique constant prend, dans le vide, un mouvement uniformément accéléré, c'est-à-dire que son accélération est constante, sa vitesse augmentant avec le temps; c'est ce qui se passe en particulier dans les tubes à vide. Au contraire, dans les mêmes conditions, les électrons libres du germanium N (ou les trous du germanium P) acquièrent un mouvement à vitesse moyenne constante, car ils sont freinés par les atomes du cristal sur lesquels ils viennent rebondir de temps en temps. La vitesse moyenne est proportionnelle à l'intensité du champ, et l'on écrira $\vec{V}$ moyen $= \mu\vec{E}$, où μ s'appelle la « mobilité ».

Pour nous représenter l'ordre de grandeur de la mobilité, disons que, dans un barreau de germanium de 1 cm de long dont les extrémités sont reliées aux électrodes d'une pile de 1,5 V, les électrons se déplacent à la vitesse moyenne de 5,3 m/s, et les trous, plus lents, à 2,5 m/s. Ce n'est pas considérable, et les électrons dans un tube à vide vont beaucoup plus vite. On a d'ailleurs longtemps pensé, pendant les premières années du transistor, que la faible valeur de la mobilité limiterait éternellement l'emploi des semi-conducteurs aux phénomènes variant relativement lentement dans le temps, aux « basses fréquences ». Mais on a su compenser ce manque de mobilité des charges par une réduction des dimensions de certains volumes dans lesquels elles se déplacent, si bien que, à l'heure actuelle, les semi-conducteurs, et en particulier les transistors, sont parfaitement aptes à fonctionner aux fréquences élevées. Le fait que la vitesse soit proportionnelle au champ signifie aussi que le courant (nombre de charges qui passent en une seconde dans l'unité de surface) est lui aussi proportionnel au champ. Et ceci est la loi d'Ohm bien connue, qui permet de définir pour les semi-conducteurs extrinsèques une « conductivité », inverse de la résistivité. Il est facile d'imaginer que la conductivité sera à la fois proportionnelle à la mobilité et à la quantité de charges libres, plus exactement au nombre de charges par unité de volume ou « concentration », soit, à la température ordinaire, au nombre d'atomes d'impureté par unité de volume.

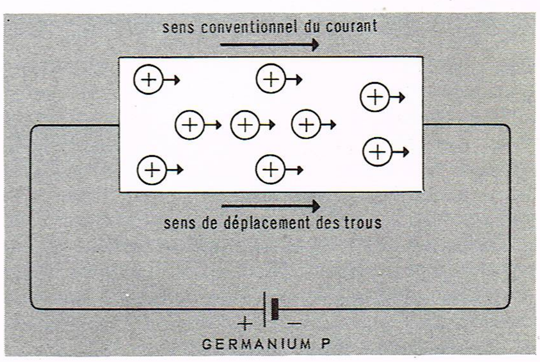
Un champ électrique a donc pour effet d'établir, dans un semi-conducteur N ou P, un courant (« de champ ») dont l'intensité est proportionnelle à la concentration en atomes d'impureté.

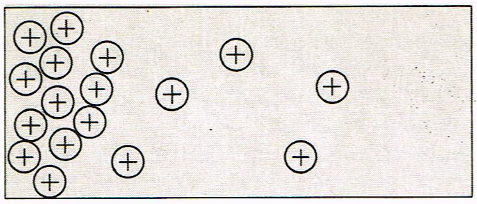
la diffusion des électrons ou des trous engendre aussi des courants

Supposons que deux bouteilles hermétiques soient réunies entre elles par un tube, fermé lui-même par un robinet, et que ces bouteilles soient remplies d'un même gaz, de l'hydrogène par exemple, mais à une pression différente. Ouvrons le robinet : la différence de pression entre les deux bouteilles aura

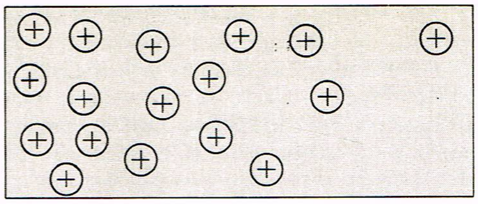


Une pile branchée aux bornes d'un barreau de germanium dopé y engendre un courant. Son sens est celui du déplacement des trous ou le sens inverse du déplacement des électrons. Que le germanium soit N ou P, le courant circule toujours du pôle + au pôle — de la pile.





Déplacement de trous par diffusion dans un barreau de germanium P. En I, la concentration en trous est beaucoup plus grande dans la partie gauche du barreau que dans la partie droite. En II, on voit la position des trous un instant plus tard ; ils se sont « étalés » dans le volume du germanium et leur mouvement y a créé un courant.



tendance à diminuer, et, au bout d'un temps suffisamment long, la pression sera la même partout; les molécules d'hydrogène contenues dans la bouteille où la pression était la plus grande ont diffusé vers l'autre bouteille. Comme, dans ce cas, la concentration (nombre de molécules par unité de volume) traduit la même réalité physique que la pression, on peut dire que la diffusion est la propriété qu'ont les molécules d'hydrogène de se déplacer des régions à forte concentration vers celles où la concentration est plus faible. Le flux de molécules sera d'ailleurs d'autant plus grand que la différence de concentration (ou de pression) sera elle-même plus importante.

Revenons aux semi-conducteurs, et supposons que l'on ait réussi, à un instant donné, à accumuler des charges (électrons ou trous) à une des extrémités d'un barreau de germanium. La diffusion traduit le fait que ces charges auront tendance à se déplacer vers l'autre extrémité. Au bout d'un temps suffisamment long, elles seraient uniformément réparties dans tout le barreau, si toutefois en se déplaçant elles ne laissaient apparaître des charges fixes, d'où un champ électrique qui agira sur leur mouvement; c'est seulement en imaginant qu'on puisse négliger ce deuxième effet qu'on pourra dire que la concentration des charges mobiles sera constante dans le barreau au bout d'un temps assez long.

En se déplaçant, les charges donnent naissance à un courant, dit « courant de diffusion ». La loi de la diffusion traduit le fait que ce courant est proportionnel au gradient de concentration des charges.

Si n (concentration dans le barreau à un instant donné) varie suivant l'abscisse x du barreau, le gradient de concentration est égal à la dérivée de n par rapport à x, et le courant de diffusion s'écrit $J = Dq\frac{dn}{dx}$, D étant une constante appelée « constante de diffusion », et q la charge de l'électron.

On devine aisément que la diffusion est liée à la mobilité; dans les semi-conducteurs où la mobilité des charges est grande, la diffusion se fera facilement. Effectivement, à une température donnée, le rapport D/μ est le même pour les trous et les électrons, et pour les différents semi-conducteurs.

L'étude de la jonction P-N va nous permettre de mieux sentir physiquement en quoi consiste la diffusion, et l'importance relative des courants de diffusion et de champ.

porteurs majoritaires, porteurs minoritaires

Dans un semi-conducteur N, la concentration en trous n'est pas rigoureusement nulle. En effet, la température libère des paires électron-trou en nombre nettement moins grand que celui des électrons libres fournis par les atomes d'impureté, mais qu'on ne peut pas toujours négliger. Les trous ainsi créés, étant en présence de nombreux électrons libres, auront tendance à se recombiner avec certains d'entre eux, donc à disparaître en tant que charges mobiles, et, finalement, leur nombre sera encore plus faible que s'ils avaient été libérés dans un semi-conducteur intrinsèque (sans impuretés). On montre d'ailleurs que le produit des concentrations en trous et en électrons libres est le même dans un semi-conducteur extrinsèque (avec impuretés) et dans un semi-conducteur intrinsèque, à température donnée. Comme dans ce dernier cas les trous et les électrons sont en nombre égal, cela s'écrit $np = n_i^2$, où n et p désignent les concentrations en électrons libres et trous du semi-conducteur extrinsèque, et n_i, la concentration en électrons (ou en trous) qu'il y aurait à la même température si le semi-conducteur était intrinsèque.

Dans un semi-conducteur N, les électrons libres sont appelés « porteurs majoritaires », et les trous, « porteurs minoritaires ». Contrairement à ce qu'on pourrait penser, ces derniers sont très importants, et tous les résultats auxquels on arrive à propos des jonctions et transistors s'obtiennent par des considérations et calculs sur les porteurs minoritaires. La relation précédente montre que la concentration en minoritaires augmente fortement avec la température. En effet, n_i augmente avec la température, et comme la concentration en majoritaires reste à peu près constante et égale à celle des atomes d'impureté, la température agit presque uniquement sur le nombre des minoritaires.

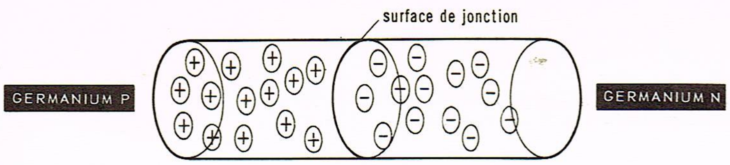
la jonction P-N, base de presque tous les semi-conducteurs

Supposons que nous puissions réaliser un contact électrique parfait, suivant une surface plane, entre un élément de germanium N et un élément de germanium P. Contact parfait signifiant que les porteurs de charge peuvent traverser la surface de jonction.

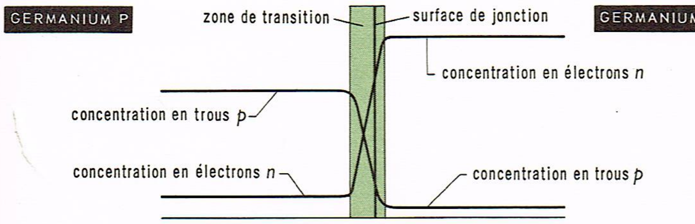
Pour l'instant, ne relions la jonction à aucune source extérieure de tension. On pourrait croire qu'il ne va rien se passer. Eh bien, au contraire: les trous de la région P qui se trouvent près de la jonction vont se précipiter dans la région N par diffusion, car, dans la région P, la concentration en trous est grande, alors que dans la région N elle est presque nulle.

De même, les électrons de la région N auront une énergique tendance à diffuser vers la région P. Mais, en quittant la région N, les électrons vont laisser derrière eux une région chargée positivement. En effet, nous avons vu que, dans un volume de semi-conducteur N, à chaque électron libre était associé un ion positif fixe (l'atome d'impureté qui a perdu l'électron en question), de sorte que l'ensemble électrons + ions est neutre électriquement; mais si les électrons désertent le volume considéré, les ions restent et sa charge globale devient positive.

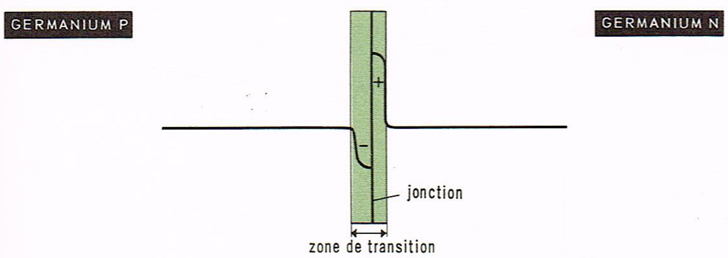
La diffusion des électrons de la région N vers la région P laisse donc apparaître une zone chargée positivement au voisinage de la jonction. Pour les mêmes raisons, il apparaît également dans la région P une zone chargée négativement au voisinage de la jonction. Ces deux zones vont créer un champ électrique, qui agira sur les trous et les électrons voisins de la jonction et aura pour effet de limiter leur pénétration dans l'autre région. Il va s'établir un équilibre, car, aucune source de tension extérieure n'étant reliée aux éléments P et N, il ne saurait y avoir de mouvement continu de charges dans un sens ou un autre.



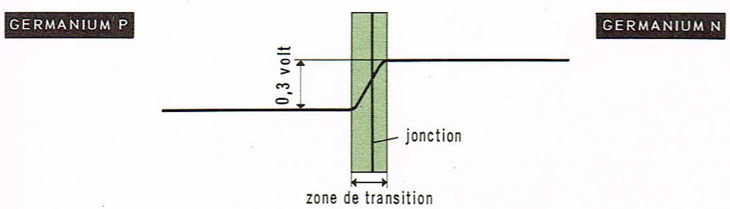
Les deux cristaux de germanium, dopés + et —, sont accolés avec précision. Ils constituent une « jonction P-N », élément fondamental des semi-conducteurs.



Concentration en trous et en électrons libres dans une jonction non polarisée. Hors de la zone de transition, très étroite (1 micron environ), les concentrations sont celles qu'il y aurait s'il n'y avait pas de jonction.

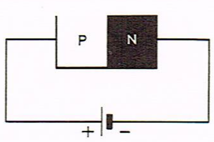


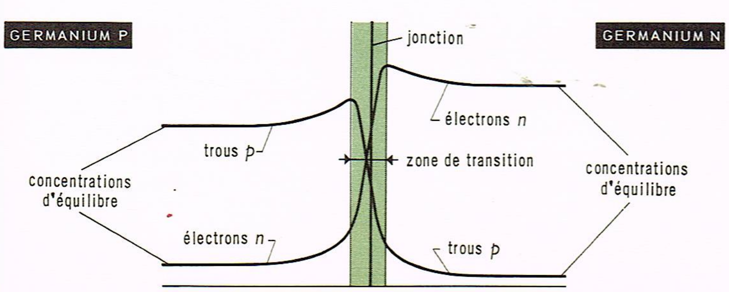
Charges électriques dans la zone de transition. En dehors de cette zone, la charge est nulle, les charges mobiles étant aussi nombreuses que les ions d'impureté de charge opposée.



Différence de potentiel de chaque côté de la jonction. Cette différence de potentiel est due aux charges qui se trouvent dans la zone de transition. Le côté P est à un potentiel inférieur à celui du côté N.

Ci-dessous, **répartition des concentrations dans une jonction polarisée en direct. Par rapport au diagramme de la jonction non polarisée, toutes les concentrations sont augmentées au voisinage de la zone de transition.**





les deux côtés ne sont pas au même potentiel

La figure ci-contre montre comment sont réparties les concentrations et les zones chargées électriquement. Le champ qui en résulte entraîne une différence de potentiel entre les régions N et P. La région P, dans une jonction non reliée à des sources extérieures de tension, possède donc un potentiel inférieur à celui de la région N. Mais il ne faut pas croire pour autant que, si l'on relie les régions N et P par un conducteur électrique, on constatera le passage d'un courant : il existe en effet des différences de potentiel entre métal et semi-conducteur qui annulent celle dont nous venons de parler.

La zone chargée électriquement s'appelle la « zone de transition ». Les calculs montrent qu'elle est extrêmement étroite, de l'ordre du micron. Ce qui signifie que, même à faible distance de la jonction, les répartitions des trous et électrons libres sont les répartitions d'équilibre, et que le champ est nul. La différence de potentiel entre les régions N et P s'appelle la « barrière de potentiel ». Elle est de 0, 25 à 0,35 V environ à la température ordinaire.

Supposons maintenant que nous appliquions une différence de potentiel entre les régions P et N, en reliant le pôle positif d'un élément de pile au côté P, et son pôle négatif au côté N.

Cette différence de potentiel aura pour effet de créer un champ électrique supplémentaire que nous admettrons localisé dans la zone de transition. Il sera dirigé de gauche à droite, et viendra donc se retrancher du champ précédent; le champ résultant, étant alors plus faible, deviendra insuffisant pour neutraliser la diffusion, et l'on assistera à un passage de trous de la région P vers la région N par diffusion, ainsi qu'à un passage d'électrons de la région N vers la région P. Le courant qui en résultera sera la somme des courants de trous et d'électrons : il sera dirigé de gauche à droite, puisque le sens conventionnel du courant (sens inverse des électrons) est celui des charges positives, c'est-à-dire celui des trous.

Après avoir traversé la jonction, les trous se trouveront dans la région N, riche en électrons libres; il y aura de nombreuses recombinaisons électron-trou, car le taux de recombinaisons est d'autant plus grand que les trous d'une part, et les électrons d'autre part, sont plus nombreux. Les trous vont donc disparaître progressivement au fur et à mesure qu'on s'éloignera de la jonction dans la région N, et, à partir d'une certaine distance de celle-ci (quelques millimètres), leur concentration sera la concentration d'équilibre des trous dans la région N, c'est-à-dire presque nulle. De même, les électrons de la région P auront presque complètement disparu quelques millimètres à gauche de la jonction. Contrairement à ce qui se passait en l'absence de polarisation, les concentrations d'équilibre ne sont atteintes qu'à des distances de la jonction bien supérieures à la largeur de la zone de transition, qui est, rappelons-le, la région où la charge électrique globale d'un élément de volume n'est pas nulle.

En dehors de la zone de transition, les courbes représentatives des concentrations de trous et d'électrons sont en première approximation parallèles. En effet, comme la charge globale est nulle dans l'élément de volume, la différence n-p doit rester constante.

on peut faire passer le courant du côté P au côté N

On peut encore faire remarquer que, dans la région N, le courant dû aux trous est un courant de diffusion, qui diminue quand on s'éloigne vers la droite. Il est remplacé progressivement par un courant d'électrons fourni par la pile, électrons qui remplacent au fur et à mesure les électrons de la région N qui se sont recombinés avec les trous. Le flux d'électrons est

dirigé de droite à gauche, le courant est donc dirigé de gauche à droite et a le même sens que le courant de trous. La somme « courant de diffusion des trous plus courant de recombinaison des électrons » est constante en chaque point de la région N. On pourrait faire un raisonnement analogue à propos des électrons de la région P.

La théorie montre que le courant global qui traverse une jonction polarisée en direct est une fonction exponentielle de la tension appliquée : il suffira d'une faible tension pour obtenir des courants importants. Pratiquement, en électronique, où les courants sont généralement inférieurs à l'ampère, la tension directe aux bornes d'une jonction sera inférieure au volt.

mais du côté N au côté P le courant ne peut pas passer

Relions maintenant le pôle positif de notre pile au côté N de la jonction, et le pôle négatif au côté P.

Le champ créé par la pile va s'ajouter cette fois à celui qui est dû aux charges de la zone de transition, et la diffusion ne pourra pas avoir lieu comme lorsque la jonction est polarisée en direct. Les porteurs majoritaires de chaque région ne peuvent donc pas traverser la jonction, et le courant est nul.

Plus exactement, il est presque nul, car les porteurs minoritaires (les trous de la région N, par exemple) voisins de la jonction sont, eux, sollicités par le champ à la traverser et produisent un courant, très faible évidemment puisqu'ils sont peu nombreux. Ce courant, appelé « courant inverse de la jonction », est de l'ordre de quelques micro-ampères pour les jonctions ordinaires au germanium, encore beaucoup moins important pour les jonctions au silicium, et dépend assez peu de la tension appliquée. En revanche, il augmente rapidement avec la température, car il est lié à la concentration des porteurs minoritaires, laquelle dépend du taux de création de paires électron-trou dans un semi-conducteur intrinsèque, c'est-à-dire de la température ($np = n_i^2$).

une trop grande tension inverse détruit la jonction

Si la tension appliquée est très importante, il peut se produire un « claquage » de la jonction. En effet, le champ dans la zone de transition devient très élevé, et communique une grande vitesse aux charges qui participent au courant inverse; celles-ci, en heurtant des atomes de germanium, leur arrachent des électrons, créant de nouvelles paires électron-trou, qui viennent augmenter le courant inverse. Comme la tension est élevée, la puissance dissipée dans la jonction le devient également; la température augmente, d'où nouvelle augmentation de courant inverse, qui conduit finalement à la destruction de la surface de jonction. (Le champ peut également arracher directement des électrons aux atomes de germanium.)

Actuellement, les jonctions supportent des tensions qui atteignent couramment plusieurs centaines de volts, et dépassent parfois 1 000 V.

Une jonction P-N présente donc les caractéristiques suivantes : elle est susceptible de laisser passer le courant lorsqu'on la polarise dans un certain sens, en présentant une faible résistance; lorsqu'elle est polarisée en sens inverse, le courant qui la traverse est limité à une valeur presque nulle, qui augmente toutefois rapidement avec la température.

C'est l'effet de diode, qui a d'innombrables applications en électronique. Les jonctions P-N, également appelées « diodes à jonction », sont ainsi des composants passifs très employés, dont nous examinerons plus loin, à propos des transistors, quelques procédés de fabrication, ainsi que les principales utilisations.

deux jonctions : un transistor

Supposons qu'un élément N soit compris entre deux éléments P formant deux jonctions avec lui; moyennant certaines conditions, l'assemblage de ces trois éléments constituera un transistor PNP. Si c'est un élément P qui est compris entre deux éléments N, il s'agira alors d'un transistor NPN. Par la suite, nous ne parlerons que des transistors PNP, les résultats qu'on obtient avec les NPN étant les mêmes, au signe des tensions près.

V. figure p. 183 en haut à droite.

Examinons la figure ci-contre, où chacun des trois éléments a été réuni par un conducteur métallique aux pôles de diverses sources de tension. Nous remarquons que la première jonction P-N, en partant de la gauche, est polarisée en direct, puisque le côté P est positif par rapport au côté N. Au contraire, la deuxième jonction est polarisée en inverse. Disons tout de suite que la première région P s'appelle l' « émetteur » du transistor, la région N la « base », et l'autre région P le « collecteur ».

Nous savons que la jonction émetteur-base va laisser passer un courant : les trous de l'émetteur vont donc pénétrer dans la base, et les électrons, de la base dans l'émetteur. Examinons ce qui va se passer si aucune précaution particulière n'est prise quant aux dimensions de la base.

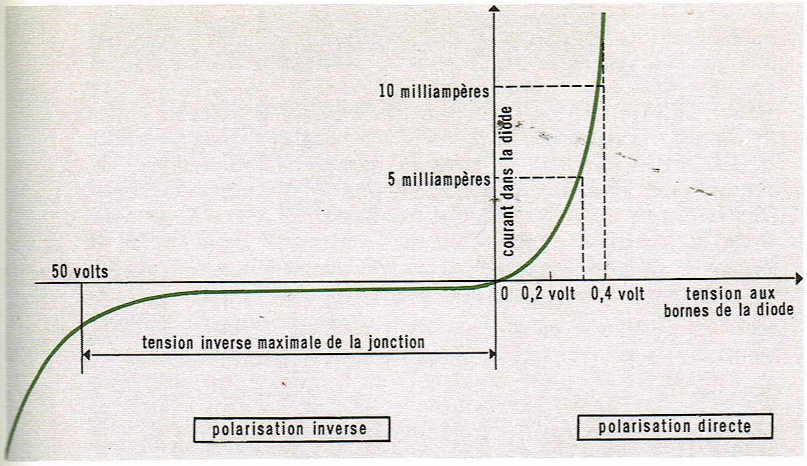
Les trous, une fois dans la base, se recombinent avec les électrons qui s'y trouvent en abondance. Comme la concentration en électrons doit y rester constante en chaque point (puisque nous sommes en régime établi), les électrons recombinés sont remplacés par d'autres électrons, provenant du conducteur métallique relié à la base. Quant aux électrons qui traversent la jonction vers la gauche, ils sont pour la même raison remplacés, au fur et à mesure qu'ils quittent la base, par d'autres électrons venant de l'électrode de base. Donc, tout le courant qui traverse la jonction émetteur-base se referme par le conducteur de base, et l'ensemble se comporte comme une jonction P-N ordinaire, la deuxième jonction ne jouant aucun rôle.

Or, pour que l'assemblage des trois éléments P, N et P soit un transistor, il faut au contraire que la plus grande partie du courant issu de l'émetteur se referme par le collecteur. Comment faire pour parvenir à ce résultat ?

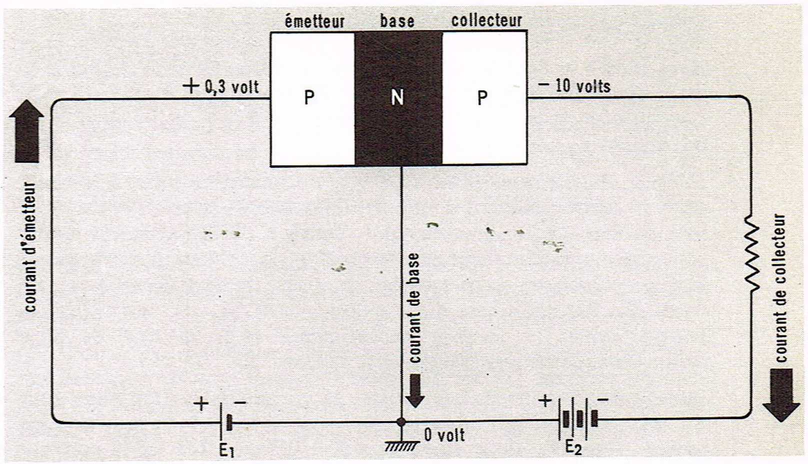
une base extrêmement mince, et nous avons un transistor

Supposons qu'on ait réussi à réaliser une base de très faible largeur (par exemple 10 à 100 μ) et que le contact entre base et collecteur soit assuré sur une grande surface. Les trous de l'émetteur qui ont traversé la première jonction (celle de gauche) se déplacent dans la base par diffusion : ils vont donc rapidement rencontrer la deuxième jonction, avant d'avoir eu le temps, pour la plupart, de se recombiner aux électrons de la base. Que se passera-t-il alors à cette jonction de droite?

Comme elle est polarisée en inverse, les trous du collecteur P ne pourront pas passer vers la gauche — cela nous l'avons vu —, ni d'ailleurs les électrons de la base N vers la droite. Mais des trous arrivant de la gauche sur cette jonction inverse seront au contraire précipités vers le collecteur par le champ qui y règne.



Courbe caractéristique d'une jonction PN.
A droite de l'axe vertical, on est en polarisation directe : une faible tension appliquée produit un courant important. A gauche de cet axe, la polarisation est inverse : même si la tension est grande, le courant reste négligeable, sauf si elle dépasse une valeur limite.



Transistor PNP polarisé normalement.
La première jonction (jonction émetteur-base) est polarisée en direct, alors que la jonction base-collecteur l'est en inverse; un courant dirigé de gauche à droite va donc traverser la première jonction. Le dispositif ci-dessus sera un transistor si la plus grande partie de ce courant, au lieu de se refermer par l'électrode connectée à la base, parvient à traverser la deuxième jonction.

un autre paramètre important : la résistivité de la base

Ainsi, les trous « émis » par l'émetteur sont presque tous « collectés » par le collecteur. Mais le courant d'émetteur comprend également le courant d'électrons qui diffuse de la base vers l'émetteur, courant qui passe par l'électrode de base. Comment le minimiser par rapport au courant des trous?

On montre que le courant d'électrons dépend de la concentration en électrons libres dans la base, de même que le courant de trous de la concentration en trous dans l'émetteur. En rendant la première beaucoup plus faible que la seconde, on obtient le résultat cherché. **Autrement dit, on fait la base avec du germanium beaucoup plus résistant que celui de l'émetteur.**

Et cela présente un autre avantage : les trous se recombinent d'autant moins vite qu'ils sont en présence d'un plus petit nombre d'électrons. Si la résistivité de la base est élevée, on sera moins exigeant en ce qui concerne sa minceur, et le transistor sera plus facile à fabriquer.

La valeur de la tension négative appliquée au collecteur est sans effet sur le fonctionnement du transistor. **Ce qui est important, c'est le fait que les trous parviennent à la jonction collecteur-base.** Là, ils seront toujours captés par le collecteur, que la tension soit de — 20 V ou de — 1 V. C'est d'ailleurs cette propriété qui permet de faire de l'amplification avec un transistor, comme nous allons le voir bientôt.

le transistor, composant actif

Avant d'examiner les différentes façons d'utiliser un transistor PNP, résumons ses principales propriétés lorsqu'il est en régime normal.

— Le courant de collecteur est presque égal au courant d'émetteur. La différence entre les deux passe dans l'électrode de base. Le rapport entre les courants de collecteur et de base s'appelle le « gain en courant » du transistor et reste à peu près constant lorsque les courants varient. Les valeurs usuelles sont comprises entre 20 et 200.

— L'émetteur est légèrement positif par rapport à la base. Comme pour les jonctions P-N ordinaires, la tension émetteur-base reste inférieure au volt pour les courants faibles, les plus fréquents.

— La tension de collecteur est négative, et peut être choisie quelconque, au moins en deçà d'une certaine limite.

où le transistor sert comme amplificateur

Et maintenant, répondons à la question suivante : comment utiliser un transistor pour en faire un amplificateur?

On pourra envoyer le signal de commande soit sur l'émetteur, soit sur la base. Quant à la puissance d'utilisation, on la prélèvera soit sur le collecteur, soit sur l'émetteur. Il existe donc plusieurs montages possibles. Le plus employé est celui à commande par la base et utilisation par le collecteur : il permet en effet la plus grande amplification de puissance et met en jeu des courants de commande faibles. C'est le montage dit « à émetteur commun ». Nous en avons donné un exemple et expliqué à son propos le mécanisme de l'amplification, sur lequel nous ne reviendrons pas.

Un autre montage (commande par l'émetteur) a été donné ci-dessus, où nous avons vu que la base est reliée au point commun des sources de tension (montage « à base commune »). Supposons que l'utilisation — la « charge » — soit constituée par la résistance placée dans l'électrode de collecteur. Si le courant de commande varie de Δi_e, le courant de collecteur, c'est-à-dire le courant d'utilisation, variera à très peu de chose près de la même quantité; le gain en courant du montage est légèrement inférieur à 1. Mais, pour faire varier le courant émetteur de Δi_e, il suffit d'une faible variation de la tension base-émetteur, qui est ici la tension de commande (par exemple 50 mV pour un Δi_e de 100 mA), alors que la tension aux bornes de la résistance de charge pourra varier de plusieurs volts; le gain en tension est donc important. Comme le gain en puissance est le produit des gains en courant et en tension, il est ici pratiquement égal au gain en tension.

le transistor en « tout ou rien »

Revenons à la figure de la page 175. Nous avons calculé qu'avec les valeurs du montage le courant de collecteur était de 5 mA. La chute de tension dans R_2 est donc de 2,5 V, et la tension de collecteur $E_2 - 2{,}5 = 2$ V.

Si nous augmentons la tension E_1, le courant de base augmente et, par conséquent, le courant de collecteur aussi. La chute de tension dans R_2 augmente, et la tension de collecteur diminue en valeur absolue. Tant que la tension de collecteur reste négative par rapport à la tension de base, le fonctionnement du transistor est celui que nous avons décrit, et, en particulier, le rapport entre le courant de collecteur et le courant de base (gain en courant) est toujours le même.

Mais, pour une certaine valeur de E_1, collecteur et base sont au même potentiel : par exemple, pour — 0,3 V, la chute de tension dans R_2 étant alors de 4,5 — 0,3 = 4,2 V, et le courant de collecteur de 4,2/500 A (8,4 mA). Que va-t-il se passer au-delà de cette valeur limite?

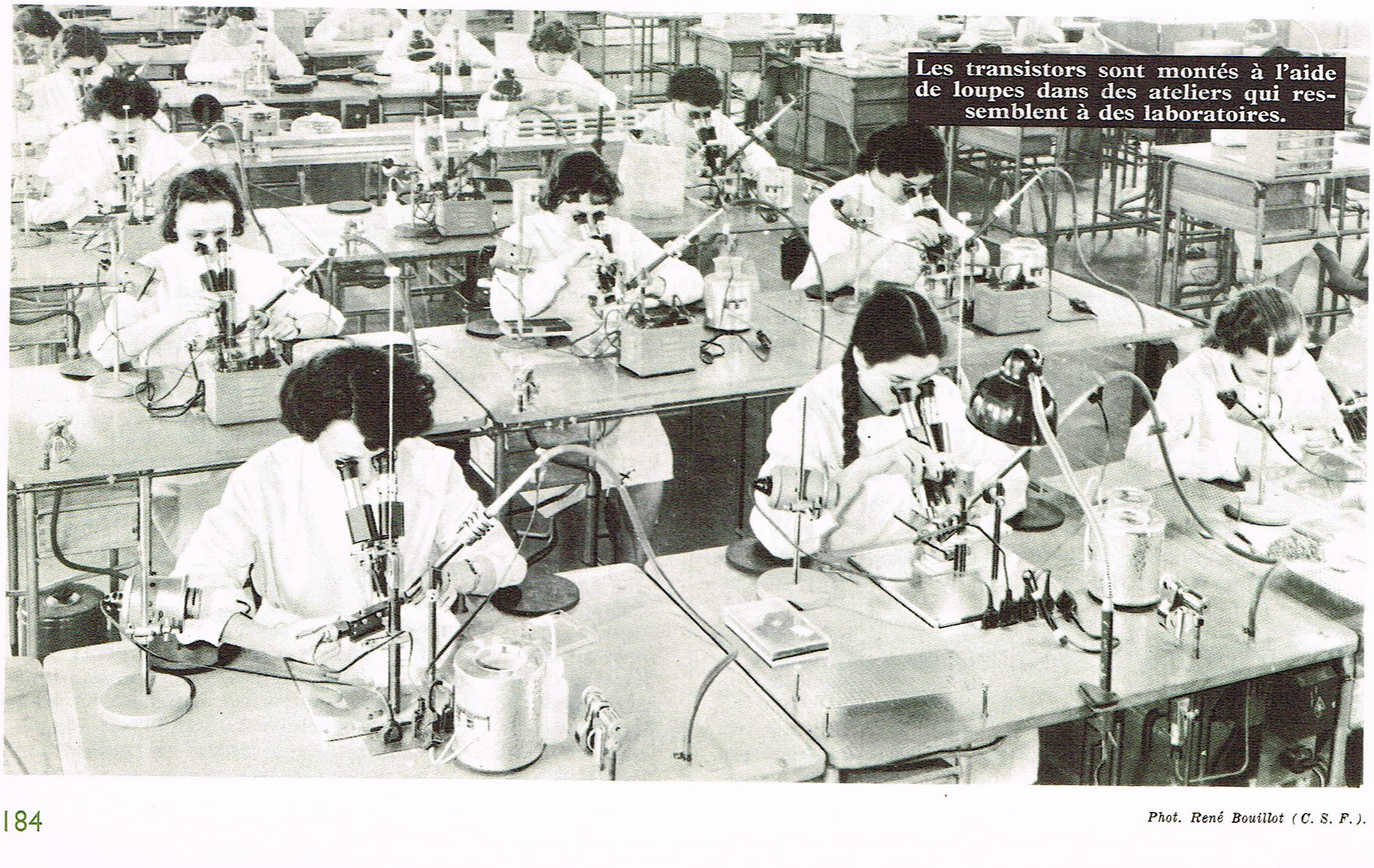
Il est clair que le courant de collecteur ne peut plus augmenter, sinon la tension de collecteur deviendrait positive par rapport à celle d'émetteur. Ce qui est impossible, le courant circulant de l'émetteur vers le collecteur. Le courant de collecteur restera donc constant; mais le courant de base continuera à augmenter proportionnellement à E_1, si bien que le rapport entre les courants de collecteur et de base, au lieu d'être égal au gain en courant du régime normal, lui sera inférieur. Le transistor est dit « saturé ». La tension de base reste faiblement négative, mais le collecteur, tout en restant négatif par rapport à l'émetteur, est positif par rapport à la base. Par exemple, sa tension sera de — 0,05 V, et la tension de la base de — 0,4 V. Toutes ces tensions sont faibles devant E_1 et E_2, et le transistor se comporte pratiquement comme si ses trois électrodes étaient en court-circuit.

On peut ainsi faire d'un transistor un interrupteur presque parfait, car la tension collecteur-émetteur est très faible une fois qu'il est saturé. Si l'on désire qu'il ne passe aucun courant dans la résistance R_2, on ramènera E_1 à zéro (ou, mieux, à une valeur positive, ce qui polarisera en inverse la jonction émetteur-base); le transistor sera alors « bloqué ». Si l'on veut que R_2 soit traversée par un courant déterminé uniquement, ou presque, par la valeur de E_2, on donnera à E_1 une valeur négative suffisante pour saturer le transistor. Ici encore, le gain entre la puissance d'utilisation, celle qui passe dans R_2, et la puissance de commande, celle qui est déterminée par E_1 et R_1, sera important.

Mais la puissance perdue — celle qui est dissipée dans le transistor — est très faible. De plus, le courant qui traverse R_2 ne dépend pas du gain en courant du transistor (ce qui n'était pas le cas en régime non saturé), et cela est un gros avantage, car ce paramètre varie fréquemment dans un rapport de 1 à 5 d'un échantillon d'un même type à l'autre.

Le transistor fontionnant en régime saturé ou bloqué — en « tout ou rien » — est susceptible d'innombrables applications : dans l'industrie, où il permet de remplacer un relais électro-mécanique avec une sécurité d'emploi accrue (pas de partie mobile sujette à l'usure); dans tous les circuits destinés au comptage d'impulsions électriques; dans les « convertisseurs », appareils permettant de transformer une tension en une autre de valeur différente avec un excellent rendement; dans les calculatrices arithmétiques, pour lesquelles il est particulièrement bien adapté (dans certaines grandes machines récentes, il entre plusieurs dizaines de milliers de transistors, presque tous utilisés de cette façon).

Au contraire, le transistor en régime normal prédomine dans les postes de radio, certains types de régulateurs et d'asservissements, les machines à calculer analogiques, les oscillateurs (circuits destinés à engendrer des oscillations sinusoïdales), et, évidemment, dans les amplificateurs de toutes sortes.



Les transistors sont montés à l'aide de loupes dans des ateliers qui ressemblent à des laboratoires.

Phot. René Bouillot (C. S. F.).

la température, origine première des limitations du transistor

Lorsqu'on utilise un transistor d'un certain type, il est bien évident qu'on ne peut pas le faire traverser par n'importe quel courant, lui faire dissiper n'importe quelle puissance, porter sa tension de collecteur à n'importe quelle valeur. Tous ces facteurs ont des limites, et c'est la température qui en est la cause principale.

D'une manière générale, tous les composants à semi-conducteurs ne peuvent fonctionner qu'entre deux valeurs limites de température.

Une valeur inférieure d'abord.

Nous avons toujours supposé qu'il y avait autant de charges libres que d'atomes d'impureté, autrement dit que ces derniers étaient tous ionisés. Or, rappelons-nous que c'est la température qui provoque cette ionisation. Si elle est trop basse, il n'y a pas assez de charges libres pour assurer un fonctionnement correct du dispositif.

Une valeur supérieure ensuite.

A température élevée, les vibrations du réseau cristallin peuvent atteindre une ampleur telle qu'elles en modifient la structure de façon irréversible. Ainsi, on ne pourra pas dépasser 100 °C avec le germanium, et 150 à 200 °C avec le silicium. Mais on devra la plupart du temps se contenter de températures plus faibles. En effet, il est nécessaire que le nombre de paires électron-trou créées par l'arrachement dû à la température reste nettement inférieur au nombre de charges libérées par les atomes d'impureté. Considérons par exemple le cas d'une diode à jonction au germanium. Son courant inverse, dû justement à ces paires parasites, sera, mettons, de 3 μA à 25 °C; mais il augmente exponentiellement avec la température : à 50 °C il atteindra 30 μA, et 300 μA à 80 °C. Si la diode est utilisée de telle sorte que son courant direct soit de 1 mA, un courant inverse seulement trois fois plus faible la rendra généralement impropre à remplir sa fonction.

Il nous reste à considérer un autre phénomène qui peut se manifester : celui de l'échauffement cumulatif.

Si la tension inverse appliquée à notre diode est assez grande, la puissance dissipée, négligeable à 25 °C, augmentera comme le courant inverse avec la température. Pour une certaine valeur de cette dernière, elle pourra devenir sensible et contribuer à l'échauffement de la zone de transition (où est localisée, comme nous l'avons admis, la différence de potentiel appliquée). Il s'ensuivra un accroissement supplémentaire du courant inverse, et, par conséquent, de la puissance dissipée, conduisant à une nouvelle augmentation de la température, qui pourra dans certains cas continuer à croître indéfiniment et provoquer la destruction de la jonction. On dit qu'il y a « emballement thermique ».

où le silicium se révèle bien supérieur

Dans le cas du silicium, les courants inverses, pour une même température, sont beaucoup plus faibles, fréquemment dans un rapport 1/100. Cela tient à ce que l'énergie qu'il faut fournir pour arracher un électron à un atome du réseau cristallin est plus élevée pour le silicium que pour le germanium. Comme le réseau lui-même supporte des températures plus élevées, on voit tout l'avantage que présentent, en ce qui concerne ce paramètre, les semi-conducteurs au silicium. Ceux-ci, bien que plus difficiles à fabriquer, donc plus chers, prennent une place de plus en plus importante, en particulier dans les équipements pour avions supersoniques, fusées, satellites, où les températures atteintes par suite du frottement de l'air sont parfois considérables.

grandeurs à ne dépasser sous aucun prétexte

Nous avons vu que la tension inverse à laquelle on peut soumettre une diode a une certaine valeur limite au-delà de laquelle le courant inverse se multiplie rapidement et provoque par échauffement la destruction de la jonction. La tension inverse de la jonction collecteur-base ne pourra donc dépasser une amplitude maximale.

En fait, lorsque le transistor est utilisé dans un montage, les choses sont un peu plus compliquées, et la tension maximale à laquelle on peut porter le collecteur est généralement inférieure.

On trouve actuellement sur le marché des transistors au germanium qui ont une tension inverse de 100 V; certains transistors au silicium peuvent supporter jusqu'à 800 V; mais les valeurs usuelles sont de quelques dizaines de volts.

Courants et puissance dissipée sont évidemment limités eux aussi dans les transistors, toujours à cause des conséquences thermiques qu'ils peuvent entraîner.

Il faut noter que, lorsqu'un transistor est traversé par un courant d'émetteur I et que la tension collecteur-base est V, la puissance VI est entièrement dissipée dans la zone de transition de la jonction collecteur-base, puisque de part et d'autre de cette zone la différence de potentiel est V. La jonction collecteur-base est donc le point chaud du transistor; comme cette surface doit être parfaite, et en particulier rester à très faible distance de la jonction émetteur-base, on conçoit les difficultés auxquelles on se heurte lorsqu'on veut faire passer de grands courants dans les transistors. Au contraire, le problème des diodes est plus facile, parce qu'il n'y a qu'une jonction, qui n'a pas besoin d'être localisée de façon aussi précise et peut être réalisée sur une plus grande surface. *On arrive à faire des diodes au silicium pour plusieurs dizaines et même plusieurs centaines d'ampères, alors que les transistors sont limités à 30 A environ pour le germanium, avec une puissance maximale de 150 W. Sur ce dernier point, les tubes restent supérieurs aux transistors, car ils sont capables d'atteindre des puissances de 100 kW.*

Mentionnons encore une autre limitation à l'emploi des transistors : ceux-ci sont sensibles à certaines radiations nucléaires, qui créent des paires électron-trou parasites comme le fait la température, et peuvent même endommager définitivement le réseau cristallin. *Cette propriété peut interdire l'utilisation des transistors dans les véhicules spatiaux, pour lesquels le rayonnement cosmique est très important. Ici encore les tubes présentent un avantage sur les transistors.*

le transistor peut-il convenir dans les cas de hautes fréquences ?

Pouvoir travailler à des fréquences de plus en plus élevées : tel a toujours été un des buts poursuivis par l'électronique. Mais qu'entend-on exactement par l'expression « hautes fréquences »?

Lorsque nous écoutons la radio, le « présentateur » nous informe de temps en temps que le poste France II, par exemple, émet sur 863 kilohertz (kHz). Cela veut dire que ce poste fabrique des courants alternatifs de forme sinusoïdale variant dans le temps 863 000 fois par seconde. Ces courants produits dans l'antenne émettrice donnent naissance à des ondes électromagnétiques de même fréquence, qui sont rayonnées dans l'espace. Nous ne nous étendrons pas sur la nature et les propriétés des ondes électromagnétiques, ni sur les problèmes posés par les antennes susceptibles de les capter, en donnant lieu à des courants de même forme que ceux du poste émetteur, mais d'intensité évidemment beaucoup plus faible (v. la Physique

des ondes, t. II). Les circuits électroniques du récepteur sont aptes à utiliser ces courants, et, en particulier, à les amplifier.

Nous savons aussi que les fréquences de la télévision et de la modulation de fréquence sont en France voisines de 200 et 90 MHz (1 MHz = 1 million de périodes par seconde). Les émetteurs et récepteurs correspondants travaillent donc à des fréquences plus de 100 fois plus grandes que ceux de la radiodiffusion. On imagine facilement qu'ils sont plus perfectionnés et que les composants dont ils sont faits, qu'il s'agisse de transistors ou de tubes, doivent posséder des qualités spéciales.

l'intérêt des hautes fréquences

Quels sont les avantages, dans le domaine des télécommunications, des hautes fréquences? Ils sont multiples.

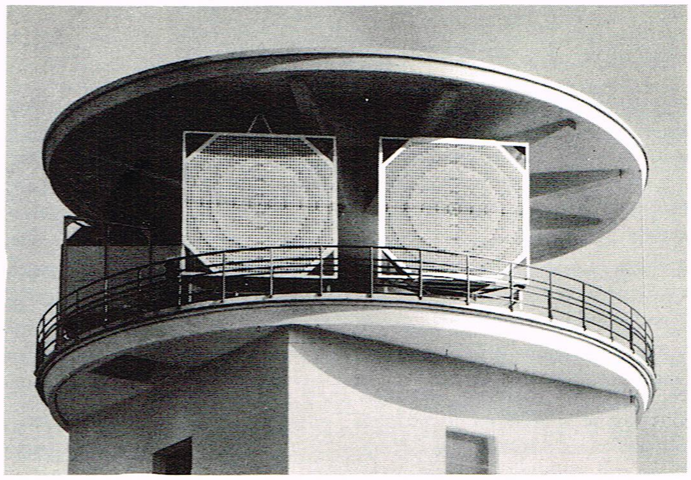
D'abord, les fréquences les plus basses sont déjà toutes utilisées. En effet, un poste émetteur n'occupe pas une fréquence unique, mais une « bande de fréquences » d'une largeur théorique de 10 kHz : si la fréquence d'émission est de 100 kHz, toutes les fréquences comprises entre 95 et 105 kHz ne pourront être utilisées par d'autres postes sous peine de brouillage; les postes les plus voisins devront émettre sur 90 et 110 kHz. Il y a déjà longtemps qu'il n'y a plus de place pour de nouveaux postes du côté des basses fréquences, et l'on doit nécessairement se tourner vers les fréquences élevées, où le champ est en principe illimité. (Rappelons que les basses fréquences correspondent aux grandes ondes et que 300 kHz équivalent à 1 000 m de longueur d'onde.) Autre avantage des hautes fréquences : la puissance qu'il faut fournir à l'émetteur est plus faible pour une même portée horizontale, d'où la possibilité de réaliser des liaisons radio plus économiques et avec des appareillages moins encombrants.

Et voici l'inconvénient : la portée est limitée, car les ondes courtes sont arrêtées par les obstacles naturels (chacun sait que, pour capter la télévision, il faut être assez près de l'émetteur et que l'antenne émettrice est placée sur un lieu élevé, genre tour Eiffel). Cet inconvénient n'en est d'ailleurs pas toujours un : un poste émettant à Bordeaux ne risquera pas d'en brouiller d'autres qui émettraient sur la même fréquence à Toulouse ou à Nantes.

Ainsi, pour quantité de liaisons radio, particulièrement lorsque l'émetteur a besoin d'être mobile, sont utilisées les hautes fréquences. Par exemple, l'aviation, civile ou militaire (aide à la navigation, approche des aérodromes), les services de police, la marine (communications dans les ports), les taxis équipés du téléphone, l'armée, etc., sont pourvus de postes à hautes fréquences.

De plus, en étudiant convenablement la forme de l'antenne, il devient possible de diriger les ondes électromagnétiques émises, avec d'autant plus de précision que la fréquence est plus grande; c'est ainsi qu'on a pu réaliser les radars, dont certains, à grande finesse d'analyse, fonctionnent à plus de 10 000 MHz. Les « câbles hertziens » utilisent également cette propriété; ils consistent en des liaisons directives entre points fixes (par exemple entre Paris et Marseille, par l'intermédiaire d'un certain nombre de relais) et permettent d'établir dans un pays un réseau de communications semblable à celui du téléphone. **Un seul câble hertzien peut véhiculer plusieurs communications en même temps, en nombre d'autant plus grand que la fréquence est plus grande; c'est le système « multiplex »,** qu'on emploie également dans les liaisons par câbles téléphoniques coaxiaux enterrés ou sous-marins. Les câbles hertziens servent aussi à transmettre les émissions de télévision et de modulation de fréquence d'un émetteur à l'autre. Les fréquences dépassent alors 1 000 MHz.

Quittons le domaine des télécommunications et des ondes électromagnétiques, que nous retrouverons bientôt, où le transistor, si transistor il y a, est utilisé généralement en régime normal, pour aborder celui où il travaille en interrupteur. Là aussi, il se pose des problèmes de rapidité. On cherche à faire des calculatrices arithmétiques capables d'effectuer de plus en plus vite les opérations qu'on leur demande. On cherche à compter des impulsions séparées par des intervalles de temps de plus en plus brefs (des besoins de cet ordre sont apparus auprès des grands accélérateurs et des piles atomiques, les impulsions électriques étant celles des détecteurs de radiations).



Relais hertzien. Les ondes venant du relais précédent sont amplifiées avant d'être dirigées sur le relais suivant.

Pour résoudre tous ces problèmes, il faut que les transistors puissent passer rapidement de l'état bloqué à l'état saturé, et inversement.

transistors lents et transistors rapides

Supposons que, dans l'émetteur d'un transistor où il ne passe aucun courant, nous injections un courant qui s'établisse de façon quasi instantanée — nous dirons un « échelon » de courant —, ce que l'on peut faire en abaissant l'interrupteur (mécanique celui-là).

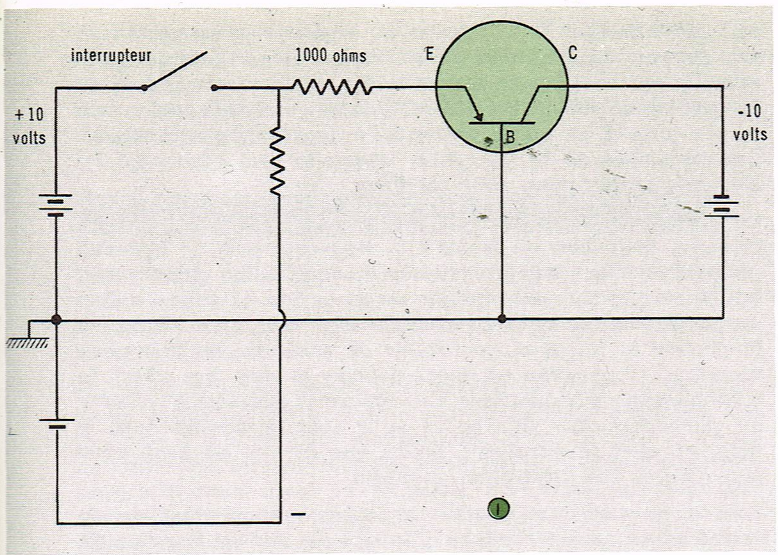
Regardons l'évolution dans le temps du courant du collecteur : il n'atteint sa valeur d'équilibre, pratiquement celle du courant d'émetteur injecté, qu'au bout d'un certain temps, variable dans de grandes proportions suivant les types de transistors.

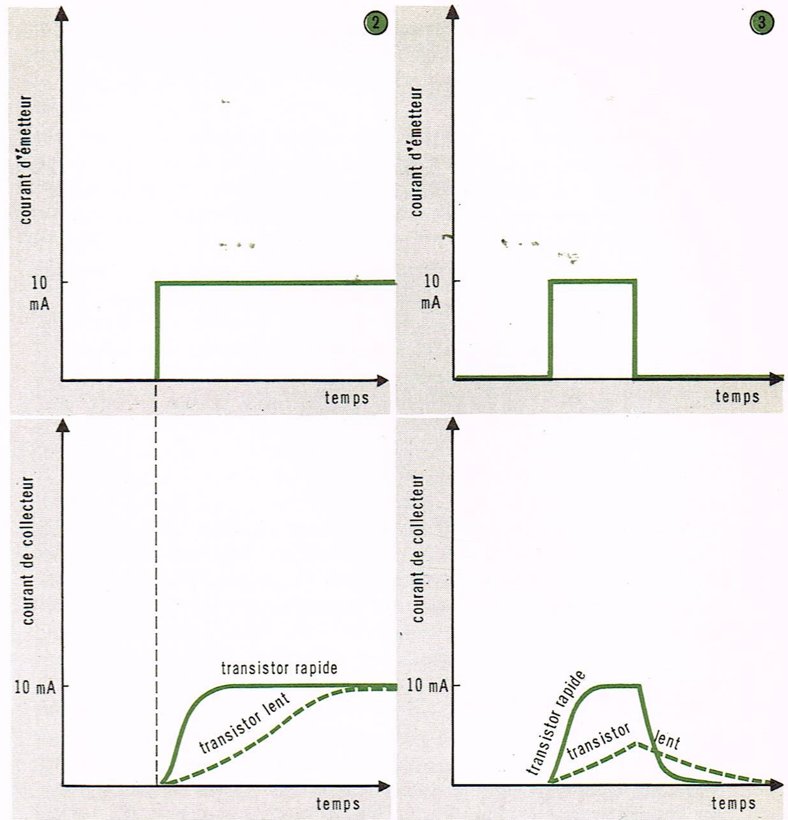
Maintenant, au lieu de l'échelon, envoyons une impulsion de courant en abaissant l'interrupteur puis en le relevant après un temps très court (en admettant que cela soit possible) : le courant du collecteur atteint ou non la valeur d'équilibre de tout à l'heure suivant que le transistor est rapide ou lent.

Supposons enfin que notre transistor soit parcouru par un courant d'émetteur moyen auquel on superpose un courant sinusoïdal de fréquence élevée. D'après ce qui précède, on conçoit facilement que, dans le cas d'un transistor suffisamment rapide, le courant du collecteur « suit » le courant d'émetteur sans le déformer : l'amplification est possible comme pour les basses fréquences. Dans le cas d'un transistor trop lent, le courant de collecteur est toujours sinusoïdal et de même fréquence, mais son amplitude maximale est plus faible et le gain de l'amplification également; au-dessus d'une certaine fréquence, celui-ci devient inférieur à 1 et le transistor cesse d'être utilisable en tant qu'amplificateur.

Quels sont les phénomènes physiques qui limitent l'emploi d'un transistor aux hautes fréquence? Il y en a plusieurs, dont le

Phot. René Bouillot (C. S. F.).





Comment un transistor « répond » à une variation infiniment rapide du courant d'entrée.

1. Le transistor est bloqué lorsque l'interrupteur est ouvert. Lorsqu'on le ferme, il s'établit dans l'émetteur, de façon instantanée, un courant de 10 mA.

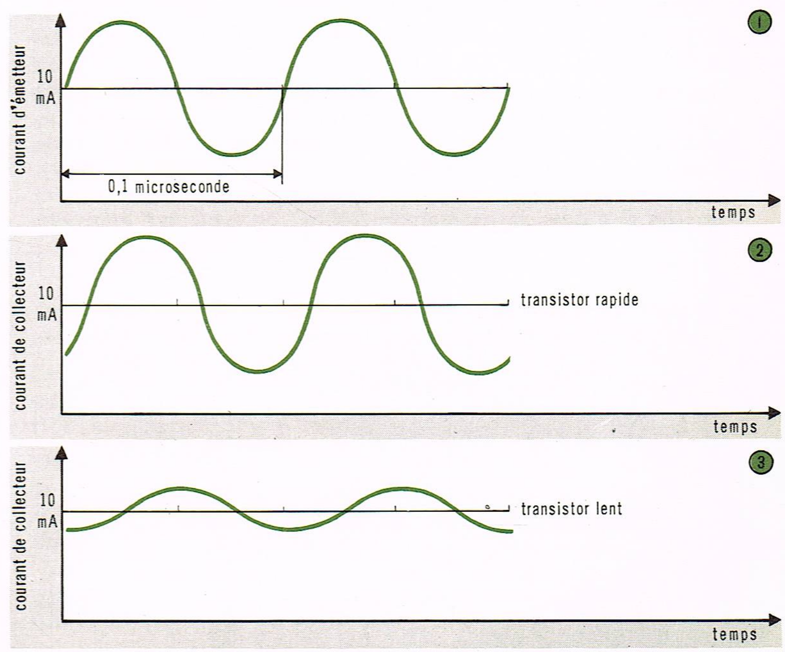
2. On voit que le courant de collecteur, qui doit, lui aussi, atteindre la valeur de 10 mA (au courant de base près), ne s'établit pas immédiatement. Le retard est plus ou moins important suivant que le transistor est plus ou moins « rapide ».

3. On a supposé qu'on fermait l'interrupteur, puis qu'on l'ouvrait de nouveau au bout d'un temps très bref. La valeur de 10 mA est atteinte ou non par le courant de collecteur suivant le type de transistor.

plus important est dû à ce qu'on appelle les « charges de diffusion ». Ce sont des charges qu'il faut amener dans la base pour qu'un courant puisse s'établir dans le collecteur, un peu comme s'il fallait charger un condensateur. On montre qu'on les réduit en diminuant la largeur de la base : **les transistors pour hautes fréquences sont donc ceux qui ont une base particulièrement mince.**

On arrive actuellement à des épaisseurs de base de l'ordre de 2μ. On devine les difficultés technologiques qu'il a fallu surmonter pour réussir à rapprocher à ce point les deux jonctions du transistor! C'est grâce à cela que les fréquences atteintes dépassent 2 000 MHz.

Les transistors sont donc parfaitement aptes à équiper les postes récepteurs de radio et de télévision. Le problème est différent pour les postes émetteurs, qui exigent de grosses puissances et restent pour l'instant le domaine des tubes à vide : tubes quasi classiques jusqu'aux alentours de 500 MHz, et tubes utilisant des principes différents au-dessus : klystrons, magnétrons, tubes à ondes progressives, etc.



Réponse du courant de collecteur à un courant d'émetteur sinusoïdal de fréquence de 10 MHz.

1 montre la forme du courant émetteur injecté : c'est un courant continu auquel on superpose un courant sinusoïdal de fréquence de 10 MHz.

2 montre la forme du courant de collecteur d'un transistor suffisamment rapide : elle est sensiblement identique à celle du courant d'émetteur.

3. Nous avons un transistor trop lent pour cette fréquence : l'amplitude maximale du courant sinusoïdal est beaucoup plus faible ; de plus, il apparaît un déphasage important par rapport au courant d'émetteur.

pour fabriquer les transistors, de sérieux problèmes

Impuretés dosées au millionième, bases épaisses de quelques microns ... on ne fabrique pas les transistors à la lime ou au chalumeau. Les techniques qu'il faut mettre en œuvre sont difficiles, onéreuses et ne peuvent être envisagées que pour des productions importantes.

De multiples procédés de fabrication ont vu le jour. Un des plus anciens, le procédé par alliage, est encore aujourd'hui le plus employé : il convient pour les transistors basse fréquence (permettant l'amplification jusqu'à quelques mégahertz) et les transistors de puissance; les autres utilisent tous plus ou moins les techniques de diffusion d'impuretés, solide ou gazeuse, qui conviennent pour les transistors haute fréquence. On les désigne par des noms étranges : mésa, mésa double diffusé, drift, épitaxial, planar, micro-alliage diffusé, laminar, et la liste n'est pas close pour autant! Car, en ce domaine comme pour tout ce qui concerne les semi-conducteurs, les techniciens font preuve d'une grande imagination; régulièrement nous arrive, trop souvent encore d'outre-Atlantique, l'annonce d'un procédé nouveau, conférant au transistor des qualités supplémentaires...

d'abord purifier

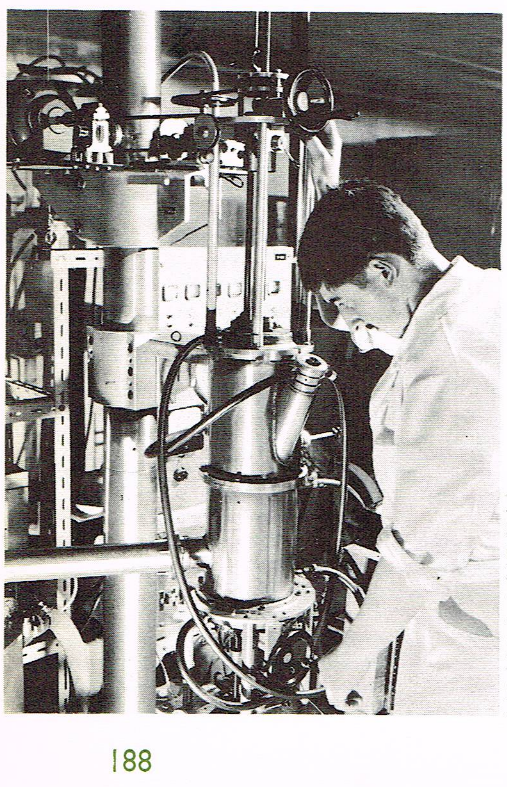
Qu'il s'agisse de faire des transistors « alliés » ou « diffusés », la première opération consiste à purifier le germanium ou le silicium. Une absence presque totale de corps étrangers est, en effet, requise à cause de la très faible proportion des impuretés N ou P qu'on ajoutera par la suite. Dans le cas du germanium, on obtient le plus souvent la purification désirée par « fusion de zone ». On part de l'oxyde de germanium, qu'on trouve dans le commerce sous forme de poudre blanche. Après réduction, purification chimique et fusion, le germanium se présente sous forme de tige à l'aspect métallique. On utilise alors la propriété que possèdent les corps étrangers dont on veut se débarrasser d'être plus solubles dans le germanium liquide que dans le germanium solide; on fond par induction une petite section de la tige, et on déplace lentement celle-ci par rapport au four : la partie fondue dissout les impuretés et les garde avec elle au cours de son déplacement le long de la tige; lorsqu'elle est arrivée à l'une des extrémités, on recommence jusqu'à ce qu'on obtienne le degré de pureté désiré. Les extrémités de la tige, dans lesquelles sont accumulés les éléments indésirables, sont sacrifiées.

Le germanium ainsi préparé ne possède pas l'arrangement cristallin désirable; on le fond de nouveau, puis on introduit un petit élément de germanium monocristallin qu'on retire ensuite très lentement : autour de ce germe, la cristallisation du germanium va s'effectuer progressivement et l'on obtiendra finalement un lingot monocristallin de quelques centimètres de diamètre. L'opération est assez délicate et doit être effectuée à température très constante. Les impuretés destinées à conférer au semi-conducteur le type N ou P sont introduites dans le bain, où elles se dissolvent, tandis que celui-ci est agité pour leur assurer une répartition uniforme.

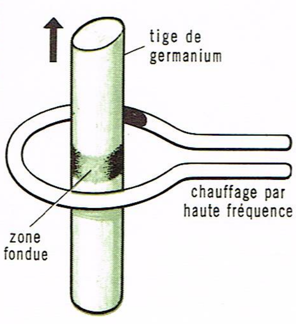
Ensuite, on détermine par des procédés optiques les plans de cristallisation, et on découpe à la scie de minces tranches de cristal parallèlement au plan convenable.

le procédé par alliage

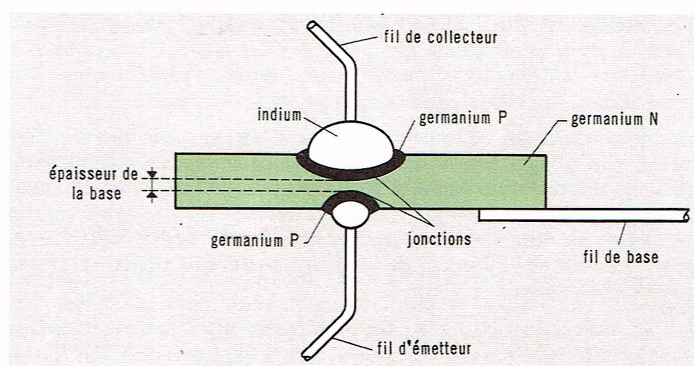
Dans le cas du procédé par alliage, le cristal est dopé avec une impureté N, et les plaquettes obtenues servent à constituer la base du transistor. Elles sont d'abord amincies par usure abrasive, puis découpées en plaquettes plus petites, dont chacune sera le point de départ d'un transistor. On réalise alors la jonction émetteur-base : pour cela, on dépose une petite bille d'indium, impureté de type P, sur la plaquette et on porte le tout à 600 °C environ, température inférieure à la température de fusion du germanium, mais supérieure à celle de l'indium. L'indium fondu dissout une certaine quantité de germanium qui, lorsqu'on refroidit, recristallise en continuité avec la plaquette, tout en emprisonnant un certain nombre d'atomes d'indium, et forme ainsi la région P de la jonction. Le reste de l'indium, une fois solidifié, constitue une excellente soudure pour le fil d'émetteur, car c'est un conducteur de l'électricité.



Préparation d'une tige de germanium pur par fusion de zone.



Les impuretés indésirables sont plus solubles dans le germanium liquide que dans le germanium solide. La zone fondue, qu'on déplace lentement le long de la tige, draine avec elle les impuretés.



Transistor PNP obtenu par alliage.

On remarquera la surface de la jonction de collecteur, supérieure à celle de l'émetteur. Le collecteur doit, en effet, capter le plus grand nombre possible des trous issus de l'émetteur. L'indium, impureté P, est déposé sur du germanium N. A 600 °C, l'indium fondu dissout une certaine quantité de germanium. Lorsqu'on refroidit, le germanium dissous recristallise en emprisonnant des atomes d'indium, constituant la partie P de la diode.

Phot. Georges Bru (C. S. F.)

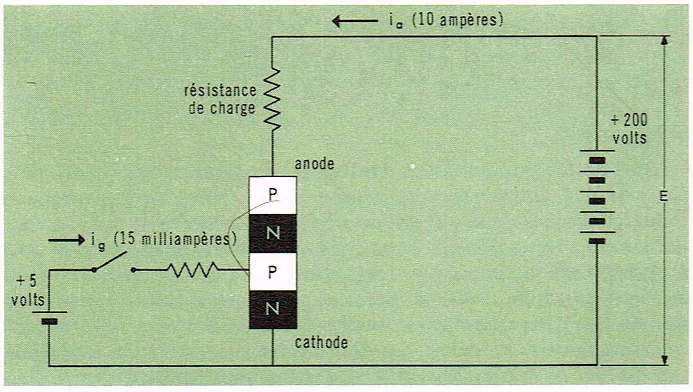
Pour faire la jonction de collecteur, il suffit de retourner la plaquette et de recommencer. Il ne reste plus qu'à souder les fils métalliques de base, d'émetteur et de collecteur, et d'enrober le tout dans un capot rigoureusement étanche à l'air. Le transistor est terminé.

les procédés par diffusion des impuretés

La technique par alliage est relativement simple et économique, mais on se heurte à de grosses difficultés lorsqu'on cherche à réduire l'épaisseur de la base pour faire des transistors haute fréquence.

On réalise alors la jonction par diffusion. On part, par exemple, d'une pastille de germanium N convenablement découpée. On la porte à une certaine température, et on l'expose à un jet de gaz inerte transportant des impuretés P. Celles-ci pénètrent progressivement dans la plaquette par diffusion, mais leur vitesse de progression, d'ailleurs très lente, est parfaitement définie par des lois connues. On arrête l'opération quand on veut : il est ainsi possible de définir le chemin parcouru par les impuretés, c'est-à-dire la position de la jonction, avec une précision de 0.2 μ. C'est la diffusion gazeuse. On utilise également la diffusion solide : les impuretés sont déposées à l'état solide sur la plaquette de germanium, dans laquelle elles pénètrent suivant des lois également bien définies à des températures idoines.

L'intérêt de ces procédés tient d'abord à ce qu'ils permettent de situer les jonctions avec une grande précision, et ensuite à ce qu'on peut obtenir de cette façon un « gradient d'impuretés » dans la base et s'arranger pour qu'il y ait plus d'impuretés du côté de la jonction d'émetteur que du côté du collecteur. Ce qui a pour effet de créer un champ électrique dans la base dirigé de telle façon qu'il accélère l'établissement du courant de collecteur, augmentant la fréquence maximale d'utilisation. Autre avantage plus important encore : la zone de transition de la jonction collecteur-base, dont la largeur augmente avec la tension appliquée sur le collecteur, peut arriver à s'étaler dans la base jusqu'à venir toucher l'émetteur, mettant celui-ci en court-circuit avec le collecteur. Or, on montre que cet étalement est d'autant moins prononcé que la base est plus fortement dopée, ce qui devient le cas lorsqu'on se rapproche de l'émetteur.



Branchement d'un redresseur contrôlé.

Lorsque l'interrupteur est ouvert, aucun courant ne passe dans le redresseur contrôlé, dont l'anode est à 200 V par rapport à la cathode. Fermons l'interrupteur : la tension anode-cathode tombe brusquement à une faible valeur, un courant important traverse le dispositif et la résistance de charge. Relevons l'interrupteur : rien ne change, le redresseur contrôlé reste dans son dernier état.

mais il n'y a pas que les transistors

En dehors des transistors, il existe de nombreux composants à semi-conducteurs. Il y a d'autres composants actifs : redresseurs contrôlés et autres structures à trois jonctions, transistors uni-jonction, transistors à effet de champ, diodes tunnel, etc. Il y a aussi des composants passifs : toute la famille des diodes, depuis la diode miniature pour « circuits » jusqu'au redresseur industriel de grande puissance. Il y a enfin des composants qui, bien que ne pouvant être classés dans aucune des catégories précédentes, jouent un grand rôle en électronique : ce sont tous ceux qui transforment en signal électrique un signal d'une autre nature. Ainsi les éléments sensibles à la lumière (cellules photo-résistantes et photo-voltaïques), aux radiations nucléaires (jonctions détectrices), à la température (thermistances), etc.

Parmi les autres composants actifs à semi-conducteurs, les plus importants, ceux dont l'avenir semble le mieux assuré, sont les redresseurs contrôlés, les diodes tunnel et les transistors à effet de champ. Sans entrer dans les détails, disons que les redresseurs contrôlés agissent un peu comme des interrupteurs : ou bien ils sont bloqués et ne laissent passer qu'un courant infime, ou bien ils laissent passer un courant important en n'opposant qu'une faible résistance. Ils sont capables de supporter de fortes tensions (jusqu'à 1 000 V) et des courants de plusieurs centaines d'ampères, ce qui les rend particulièrement bien adaptés aux besoins de l'industrie : commandes de moteurs, systèmes d'alarme, de régulation, etc.

Quant aux diodes tunnel, inventées par le Japonais Esaki en 1957, elles ont trouvé d'intéressantes applications dans le domaine des hautes fréquences, où elles permettent d'aller nettement plus loin que les transistors les plus rapides.

à quoi servent les jonctions P-N

Tout élément qui, comme une jonction P-N, laisse passer le courant dans un sens et non dans l'autre est une diode.

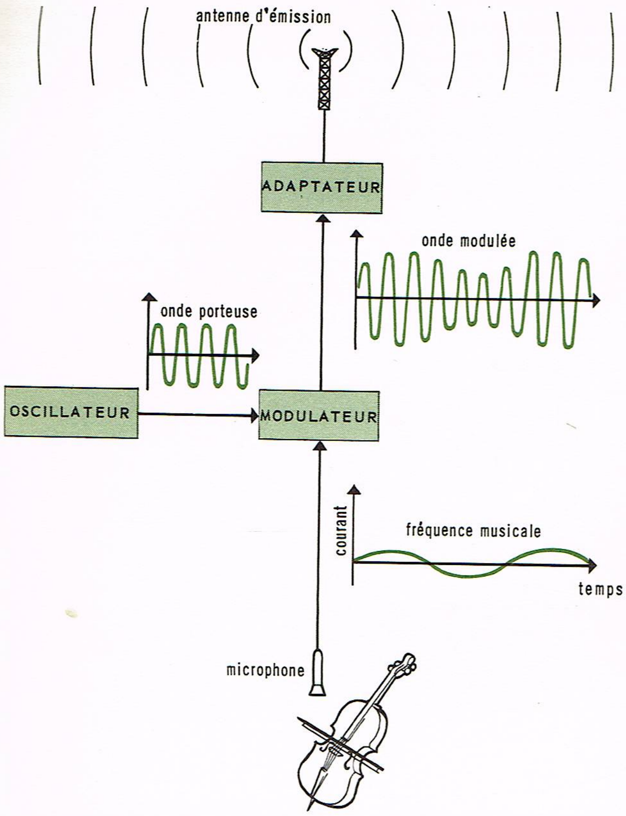
La fonction diode est fondamentale en électronique, et particulièrement en radio, car elle permet la « détection », opération de base de la partie réception.

Ouvrons ici une parenthèse pour rappeler brièvement comment on peut transmettre un morceau de musique exécuté dans les studios du poste émetteur jusqu'au haut-parleur d'un poste récepteur situé à quelques centaines de kilomètres.

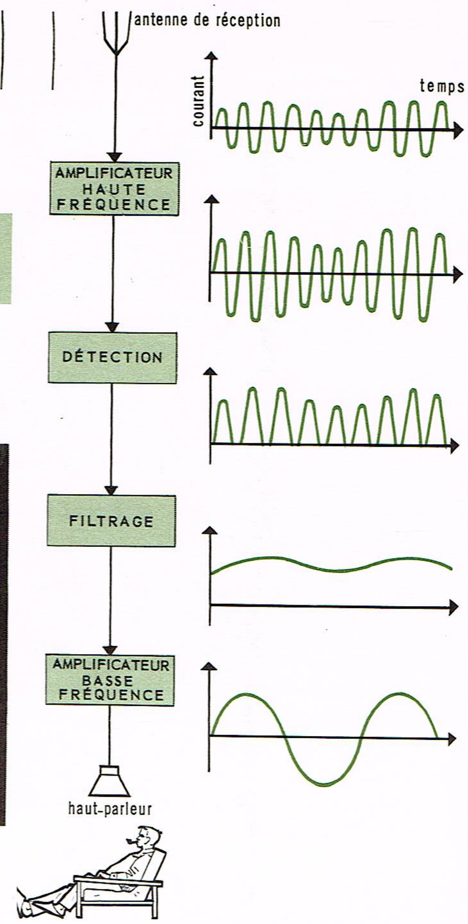
la radio, première application de l'effet diode

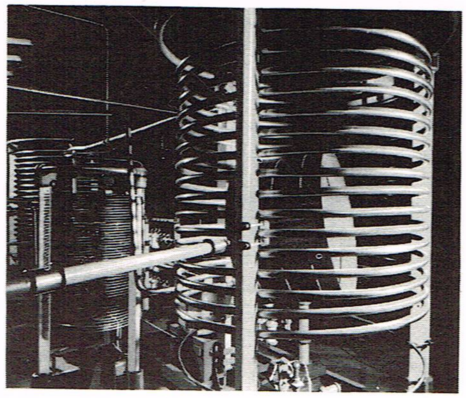
Un son musical consiste en des variations périodiques de la pression de l'air, de fréquence comprise entre 60 et 15 000 cycles par seconde, la fréquence étant caractéristique de la hauteur du son. La parole correspond également à des variations à peu près périodiques comprises dans la même gamme de fréquences. Le principe de l'émission revient à envoyer dans l'antenne un courant sinusoïdal de fréquence beaucoup plus élevée, « modulé » par les fréquences musicales, c'est-à-dire dont l'enveloppe reproduit la forme des variations de pression.

Pour cela, on commence par transformer en signal électrique de même forme les variations de pression : c'est le rôle du micro. On fabrique d'autre part le courant sinusoïdal de haute fréquence (qui définit la fréquence, ou la longueur d'onde, du poste émetteur) avec un oscillateur. On module ce courant

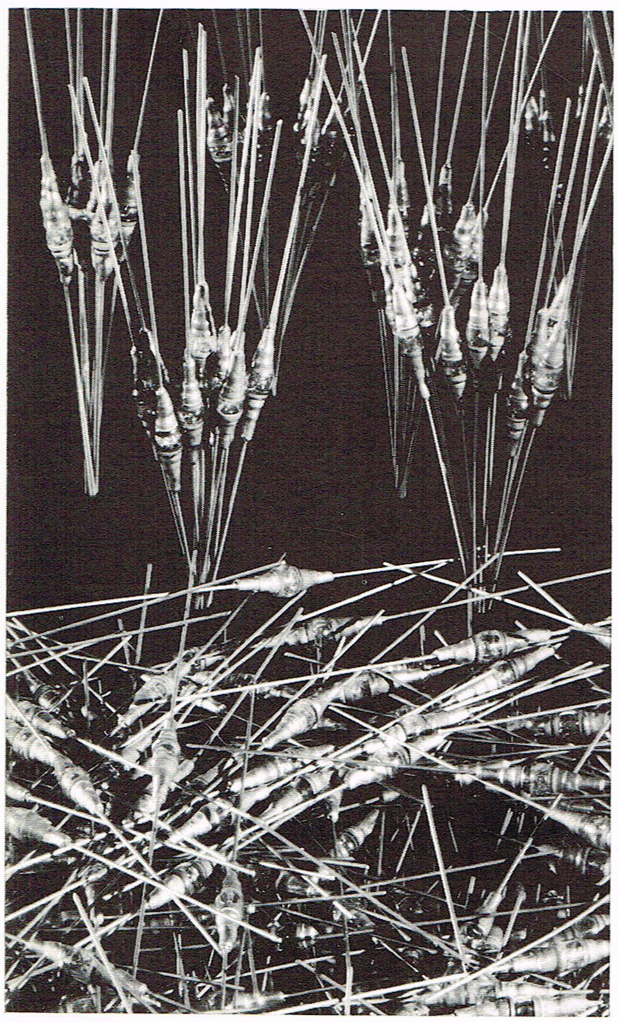


Principe de l'émission et de la réception des sons par ondes électromagnétiques.





Circuits de couplage de l'émetteur d'Allouis.



Diodes au germanium.

par le courant issu du micro dans un circuit appelé « modulateur » et le courant obtenu est envoyé dans l'antenne.

A la réception, on obtient, après amplification, un courant alternatif semblable à celui qui sort du modulateur. Le problème consiste à refaire les opérations du poste émetteur, mais en sens inverse. C'est ici qu'intervient la détection : avec une diode, on supprime toute la partie négative du courant, puis on filtre avec un filtre « passe-bas » (circuit passif simple qui laisse passer les basses fréquences et non les hautes) : il reste alors un courant qui est une reproduction fidèle de celui qui est délivré par le micro. Envoyé dans le haut-parleur, la membrane de celui-ci, en vibrant à la même fréquence, restitue le son initial.

Aux premiers temps de la radio, le récepteur était beaucoup plus simple, puisqu'il n'y avait aucune amplification, et se réduisait à la partie détection. Celle-ci, détail piquant, se faisait avec un semi-conducteur; il ne s'agissait évidemment pas de jonction P-N, mais on avait constaté que certains cristaux, en particulier la célèbre galène, jouissaient de la propriété des diodes lorsqu'on appliquait sur leur surface une pointe métallique. Par la suite, on abandonna les cristaux au profit des diodes et triodes à vide, et c'est seulement beaucoup plus tard qu'on est revenu aux semi-conducteurs.

Même à l'heure actuelle, toutes les diodes ne sont pas des jonctions classiques, et bon nombre de diodes pour circuits, c'est-à-dire pour courants faibles, sont à pointe. Le semi-conducteur est du germanium ou du silicium N et constitue la cathode, le courant passant de la pointe métallique au cristal. La théorie de ce type de diode n'est pas encore parfaitement connue : on admet qu'il se forme dans le semi-conducteur un très petit volume P au voisinage de la pointe.

Phot. Bouillot (C. S. F.).

le redressement du courant alternatif

Très importantes en radio et en électronique, les diodes ne le sont pas moins en électrotechnique, car elles permettent de « redresser » le courant alternatif, c'est-à-dire de le transformer en courant continu. A ces diodes qui travaillent à forts courants et doivent être capables de supporter des tensions inverses élevées, on donne plus particulièrement le nom de « redresseur ». Actuellement, la plupart des redresseurs sont constitués de jonctions P-N au germanium et surtout au silicium, qui peut supporter des températures plus élevées; avec ce dernier on obtient des tensions inverses dépassant 1 000 V et des courants directs de plusieurs centaines d'ampères.

d'autres applications plus inattendues

Les jonctions P-N ont d'autres utilisations que celles, pourtant fort nombreuses, qui font intervenir leur propriété de diode. Certaines d'entre elles, conçues à cet effet, sont employées comme régulateur de tension. La tension régulée est la tension de claquage de la diode, qui est dans ce cas rendue particulièrement franche, de sorte qu'elle dépend très peu du courant (inverse mais cette fois important) qui la traverse. Le claquage d'une jonction n'est pas, en effet, un phénomène destructif, pourvu qu'on prenne le soin de limiter le courant inverse de façon à ne pas dépasser la puissance maximale admissible. De tels dispositifs s'appellent « diodes Zener ».

D'autres jonctions sont utilisées comme capacités variables avec la tension. Polarisée en inverse, une diode à jonction présente une capacité due à la zone de transition chargée qui entoure la jonction. On montre que cette capacité varie avec la tension inverse appliquée. Cette propriété a trouvé de nombreuses applications dans le domaine des très hautes fréquences (amplificateurs et oscillateurs « paramétriques »).

comment détecter des radiations nucléaires

Voici encore une application récente des jonctions P-N : elles peuvent servir à détecter des radiations nucléaires émises par les corps radio-actifs, les désintégrations, le rayonnement cosmique. Celles-ci ont, sur un cristal de semi-conducteur, le même effet que la température : elles libèrent des paires électron-trou en arrachant des électrons périphériques aux atomes du cristal. Pour déceler les charges créées, voici comment on procède. La jonction est polarisée en inverse de façon à ne laisser passer qu'un très petit courant (on utilise des jonctions au silicium dont les courants inverses sont 100 à 1 000 fois plus faibles que ceux des jonctions au germanium). Si les électrons et les trous apparaissent loin de la jonction, ils auront le temps de se recombiner avant d'atteindre celle-ci, qui ne sera parcourue par aucun courant supplémentaire, et nul effet ne sera décelable. Si, au contraire, ils ont été émis suffisamment près de la zone de transition pour qu'une bonne proportion d'entre eux y parviennent, les porteurs minoritaires (les électrons si nous sommes dans la région P) se feront happer par le champ électrique qui y règne et, en traversant la jonction, donneront naissance à une impulsion de courant, venant s'ajouter au courant inverse normal. Cette impulsion est suffisamment importante pour qu'on puisse l'amplifier et, par exemple, la compter ou mesurer son amplitude dans des appareils électroniques adéquats.

En pratique, une jonction détectrice est réalisée à partir d'un cristal de silicium P à faible taux d'impuretés, et d'une partie N, exposée aux radiations, très étroite et à taux d'impuretés nettement plus grand. Dans ces conditions, la zone de transition s'étend surtout dans la partie P et son épaisseur (qu'on rend aussi grande que possible en augmentant la tension inverse) peut être beaucoup plus importante que celle de toute la région N. Ainsi les radiations à détecter effectueront une grande partie de leur parcours dans la zone de transition, où la totalité des trous et des électrons émis sont collectés par le circuit. On arrive, grâce aux techniques de diffusion, à placer la jonction à moins de 10 μ de la surface détectrice, alors que la zone de transition s'étend sur plusieurs centaines de microns dans la région P.

semi-conducteurs et lumière

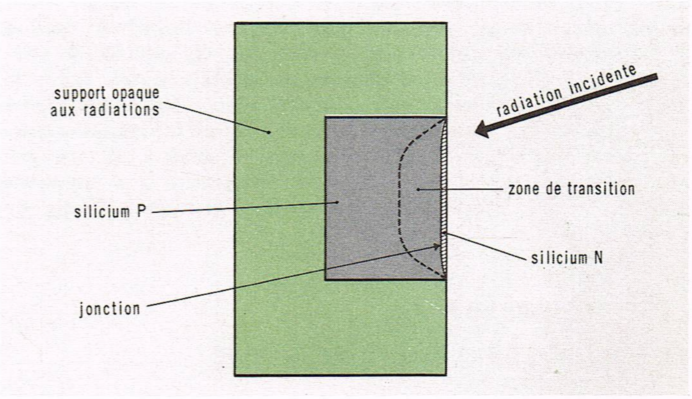
Cellule photo-électrique : ce nom nous est familier, et nous en connaissons quelques applications, telles que les posemètres des appareils photographiques, l'ouverture automatique des portes, le comptage d'objets, etc.

Il faut, en fait, distinguer deux grandes catégories : la première comprend les éléments dont la résistance diminue lorsqu'ils sont éclairés, ce sont les cellules photoconductrices; la deuxième comprend ceux qui transforment directement l'énergie lumineuse en énergie électrique : on les appelle cellules photovoltaïques.

pour détecter la lumière : les cellules photoconductrices

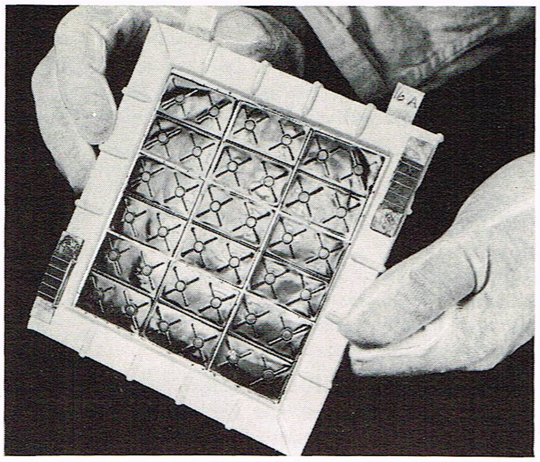
Comme les radiations nucléaires, les radiations lumineuses produisent dans un cristal de germanium ou de silicium des paires électron-trou, qu'on pourra détecter en polarisant une jonction P-N en inverse, pourvu que la lumière éclaire une zone située près de la jonction. Le courant augmente lorsque la cellule est éclairée : nous avons bien là une cellule photoconductrice.

On peut augmenter la sensibilité en utilisant non plus une diode, mais un transistor dont la base est débranchée. En envoyant la lumière sur la surface de la base, on bénéficie de l'amplification du transistor et le courant recueilli est plus grand. On réalise même des dispositifs plus complexes fondés toujours sur le même principe : par exemple, des redresseurs contrôlés qu'un pinceau lumineux suffit à faire basculer de l'état bloqué à l'état conducteur.



Jonction détectrice de radiations nucléaires.

Les paires électron-trou créées par une radiation nucléaire le long de son trajet donnent naissance à une impulsion de courant si elles sont émises dans la zone de transition ou en son voisinage immédiat. Pour obtenir une impulsion suffisante, on polarise la jonction avec une forte tension inverse, de façon à rendre la zone de transition aussi profonde que possible, et on place la jonction à très faible distance de la surface exposée. Dans ces conditions, une particule incidente effectue la plus grande partie de son trajet dans la zone de transition.



Batterie solaire.

Mais bien avant qu'on connaisse les jonctions P-N, on savait faire des éléments photoconducteurs avec certains cristaux semi-conducteurs tels que des sulfures, des oxydes, additionnés d'impuretés plus ou moins bien définies. On n'en connaissait guère la théorie, et maintenant encore celle-ci n'est pas complètement élucidée. Dans ces cristaux, les recombinaisons des trous et électrons créés par la lumière se font moins vite que dans le germanium ou le silicium, à cause de la présence de « pièges », atomes d'impureté qui fixent pendant un certain temps un des types de porteurs et laissent à l'autre la possibilité de traverser le cristal et d'augmenter sa conductivité. Ce type de cellule, encore assez employé, tend néanmoins à être remplacé progressivement par les diodes.

pour créer de l'électricité à partir de la lumière : les cellules photovoltaïques

Pouvoir transformer directement l'énergie solaire en énergie électrique ... si on arrive un jour à le faire dans de bonnes conditions de prix et de rendement, que de changements en vue sur notre planète!

Mais les résultats obtenus sont déjà encourageants. Une cellule photovoltaïque, ou « pile solaire », est actuellement une diode au silicium à grande surface de jonction, qui constitue une source de courant lorsqu'une région voisine de la jonction est éclairée.

Le principe est de même nature que celui des jonctions détectrices de radiations ou photoconductrices. Supposons qu'on éclaire une partie de la région P proche de la jonction. Les trous et électrons créés par la lumière vont se déplacer par diffusion dans le cristal et un certain nombre atteindront la zone de transition. Dans celle-ci, même lorsque la jonction n'est pas polarisée, règne un champ électrique dirigé de la région N vers la région P : ce champ arrêtera le mouvement des trous, mais attirera les électrons et les fera passer dans la région N. Si les deux électrodes de la diode sont réunies extérieurement par un circuit électrique, on constatera qu'il y passe un courant dirigé d'après le sens conventionnel de l'électrode P à l'électrode N. Si c'est la région N qui est éclairée, ce sont les trous créés qui vont traverser la jonction et donner lieu à un courant, ayant toujours le même sens que le courant d'électrons précédent.

Une des applications les plus spectaculaires des cellules photovoltaïques au silicium est l'emploi qui en est fait dans les satellites artificiels et dans les engins interplanétaires, où elles fournissent l'énergie électrique nécessaire au fonctionnement des appareils électroniques.

les tubes électroniques, ancêtres encore utilisés

Vers 1950, le chapitre relatif à l'électronique aurait commencé ici... Il n'est pourtant pas exagéré d'affirmer qu'il y a aujourd'hui deux fois plus à dire sur ce sujet qu'à cette époque, qui apparaît déjà comme un lointain passé. Dans le choix indispensable que nous avons dû faire, nous avons donné la préférence aux semi-conducteurs, et nous ne ferons qu'esquisser les propriétés des tubes, d'ailleurs plus faciles à comprendre que celles des transistors. Qui l'eût imaginé il y a quelques années, quand les tubes constituaient à eux seuls l'âme de l'électronique?

l'effet thermo-ionique, principe de base des tubes à vide

Alors que dans les semi-conducteurs les électrons participant au courant restent à l'intérieur du cristal, dans les tubes ils se propagent dans le vide ou dans un gaz. Il faut donc que, quelque part, ils puissent s'échapper du métal qui les véhicule : c'est ce que permet l'effet qu'on appelle assez improprement « thermo-ionique » (il vaudrait mieux dire thermo-électronique). Ce phénomène sera étudié au tome II. Disons seulement ici qu'il consiste en ceci : les corps conducteurs émettent des électrons vers l'extérieur lorsqu'ils sont portés à température élevée. La température provoque une agitation des électrons, qui acquièrent une énergie pouvant être suffisante pour leur faire vaincre les forces qui les retiennent à l'intérieur du corps solide. La quantité d'électrons émis dépend évidemment de l'énergie fournie aux électrons — c'est-à-dire de la température — et de l'importance des forces de liaison, qui varient suivant la nature des corps.

la diode à vide

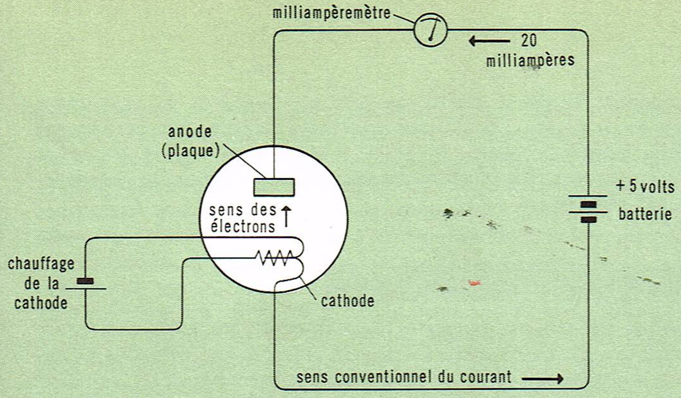
Imaginons l'expérience suivante, pour mettre en évidence l'effet thermo-ionique. Dans une ampoule, faisons arriver un fil conducteur (cathode) et un dispositif permettant de le chauffer localement (résistance chauffante l'entourant). Toujours dans l'ampoule fixons une plaque métallique (anode) et portons-la, après avoir fait le vide, à un potentiel positif par rapport à la cathode, au moyen d'une batterie extérieure.

Si la cathode est chauffée assez vigoureusement, on constate le passage d'un courant dans le circuit. Ce qui se passe est simple à comprendre : les électrons qui peuvent s'échapper de la cathode par effet thermo-ionique sont attirés vers la plaque à cause de son potentiel positif. Rien ne s'oppose à leur déplacement, puisqu'on a pris la précaution de faire le vide. Ils atteignent la plaque cependant que d'autres électrons les remplacent dans la cathode.

Si maintenant nous inversons les fils de la batterie, le potentiel négatif de la plaque repousse les électrons thermo-ioniques, qui retombent sur la cathode : aucun courant ne circule dans le circuit extérieur.

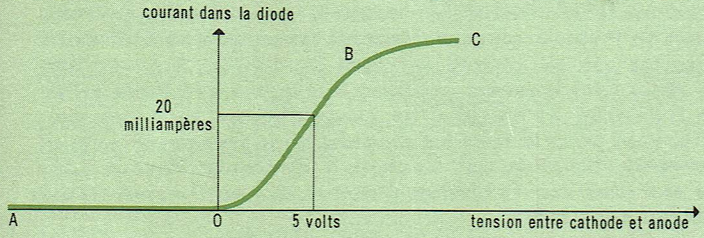
Nous avons donc là une diode : c'est la diode à vide, qui constitue la base de tous les autres tubes. Sa découverte et

Phot. U. S. I. S.



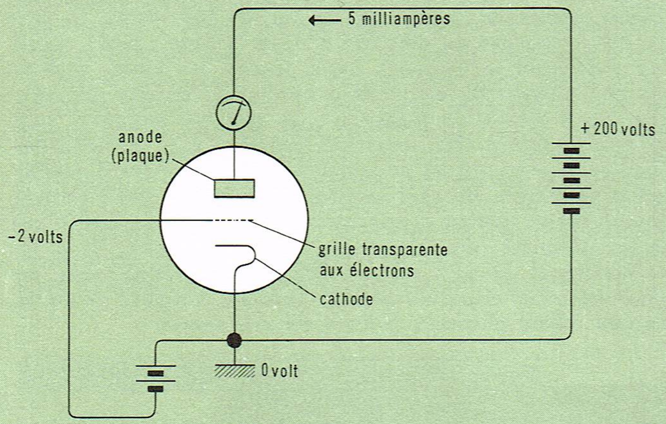
Principe de la diode à vide.

La cathode, chauffée électriquement par un circuit indépendant, émet des électrons par effet thermo-ionique. Si l'anode est portée à un potentiel positif par rapport à celui de la cathode *(cas ci-dessus)*, elle attirera les électrons et un courant passera. Dans le cas contraire, les électrons thermo-ioniques seront repoussés et retomberont sur la cathode, ne laissant passer aucun courant.



Caractéristique d'une diode à vide.

La partie OA correspond à la polarisation inverse ; seul un courant extrêmement faible peut passer. Dans la partie OB, la diode est polarisée dans le sens direct ; on remarque cependant que, dans ce cas, sa résistance est loin d'être négligeable, et qu'il faut appliquer une tension importante entre anode et cathode pour obtenir un courant appréciable. Cette résistance directe est due à ce qu'on appelle la « charge d'espace » : lorsque le courant passe, tout le volume compris entre l'anode et la cathode est constamment rempli d'électrons, formant une charge négative qui s'oppose au déplacement des électrons émis et en limite automatiquement le flux. Enfin, dans la partie BC, la tension est suffisante pour que la charge d'espace n'intervienne plus, et le courant est limité par le débit maximal d'électrons que peut fournir la cathode par effet thermo-ionique.



Branchement d'une triode à vide.

La plaque est fortement positive par rapport à la cathode. Sans la grille, un fort flux d'électrons passerait dans le tube, mais cette électrode — qui est légèrement négative par rapport à la cathode — repousse une partie des électrons et en limite le flux. En faisant varier le potentiel de grille, on fait varier aussi proportionnellement le courant dans le circuit anode-cathode.

son emploi en tant que détecteur (1904) ont fait faire un important progrès à la radio-électricité et ont permis la mise au point, quelques années plus tard, de la triode, qui a marqué la véritable naissance de l'électronique.

On a représenté ci-contre la « caractéristique » d'une diode à vide. On se reportera à celle d'une diode à jonction et l'on remarquera la différence sensible entre les deux courbes. En particulier, la tension directe nécessaire pour obtenir un courant donné est beaucoup plus faible pour la diode à semi-conducteur, ce qui est un avantage pour cette dernière. Par contre, les courants inverses des diodes à vide sont nettement plus petits que ceux des jonctions classiques.

Le chauffage de la cathode est le plus souvent « indirect ». Celle-ci a généralement la forme d'un cylindre creux, à l'intérieur duquel on place la résistance chauffante. Elle est constituée d'oxydes alcalino-terreux qui ont la propriété d'émettre des électrons en grande quantité à température relativement basse.

une diode à vide + une grille = une triode

La triode est une diode à vide à laquelle on a ajouté une troisième électrode, la « grille », qui, comme le suggère son nom, consiste en un réseau de fils métalliques placé entre anode et cathode.

La plaque étant fortement positive, on porte la grille à un potentiel négatif par rapport à celui de la cathode, de façon à établir un champ électrique qui repoussera les électrons thermo-ioniques attirés par la plaque. Si le potentiel est assez négatif (— 10 V), le champ est suffisant pour repousser tous les électrons, le courant entre plaque et cathode est nul. Si on diminue le potentiel de la grille, le champ, moins important, devient insuffisant pour contenir tous les électrons et certains d'entre eux réussissent à traverser la grille, dont les mailles leur permettent de passer, et se précipitent sur la plaque : un courant passe. Si on continue à réduire la tension de la grille, la proportion d'électrons qui parvient à la traverser augmente, le courant devient de plus en plus grand.

Pour faire varier le courant dans la triode, on dispose donc

Triode de haute puissance (15 kW) refroidie par circulation d'air forcé.



Phot. Philips.

de deux moyens. On peut soit faire varier la tension de la plaque comme pour la diode, soit agir sur celle de la grille. Mais la deuxième solution présente l'avantage de permettre une amplification de puissance considérable : le courant qui passe dans la grille est, en effet, presque nul tant qu'elle reste négative par rapport à la cathode (ordre de grandeur : 10^{-10} ampères, soit un dix-millième de microampère), si bien que la puissance d'entrée nécessaire pour modifier le potentiel de la grille est aussi presque nulle. Au contraire, on peut obtenir une puissance de sortie appréciable aux bornes d'une résistance placée dans le circuit de plaque ou dans celui de cathode.

Pour pallier certains défauts de la triode, notamment aux fréquences élevées, on a été amené à rajouter une, puis deux grilles supplémentaires. Le tube à trois grilles, ou penthode, a connu un gros développement. Jusqu'à ces dernières années, c'était le composant actif le plus répandu en électronique.

tubes ou transistors ?

Si le transistor tend aujourd'hui à remplacer le tube classique, ce n'est pas parce qu'il présente sur ce dernier des avantages décisifs sur le plan des performances; c'est même souvent le contraire, tout au moins à l'heure actuelle : le tube permet en particulier d'amplifier des signaux plus faibles, et présente un meilleur compromis fréquence-puissance, c'est-à-dire qu'il peut travailler à fréquence relativement élevée, tout en dissipant une forte puissance (les émetteurs de radio et de télévision emploient des tubes). Mais le transistor l'emporte sans conteste en ce qui concerne la commodité d'emploi et la sécurité de fonctionnement.

commodité d'emploi

Elle est évidente. Le transistor est beaucoup moins encombrant, fonctionne avec des tensions au moins dix fois plus faibles, ce qui permet de l'alimenter par des batteries ou des piles ordinaires, n'a pas besoin de chauffage de cathode, gros consommateur de puissance, reste nécessairement à des températures voisines, dans la plupart des cas, de la température ambiante, alors que les tubes sont souvent à plus de 100 °C, ce qui ne peut que nuire aux éléments situés dans leur voisinage. De plus, on ne retrouve pas dans les tubes cette aptitude des transistors à fonctionner en interrupteurs : il n'est pas possible, en effet, de réduire la tension plaque-cathode à une valeur très faible, comme c'est le cas de la tension collecteur-émetteur d'un transistor saturé, et la puissance dissipée reste toujours importante. Petitesse et faible consommation ont permis de réaliser avec des transistors des équipements qui auraient été impensables avec des tubes : citons leur emploi dans les satellites artificiels, où l'énergie électrique fournie par des piles solaires est très strictement limitée, ainsi que le poids et le volume disponible.

On se rendra compte du progrès que présente le transistor quand on saura qu'il y a dans Telstar, satellite de 80 cm de diamètre, 2 800 semi-conducteurs! Et ce nombre apparaîtra vraisemblablement dérisoire dans quelques années...

sécurité de fonctionnement

Elle ne constitue pas, on s'en doute, un avantage de moindre importance. Les tubes ont, par nature, une « durée de vie » limitée, due à la cathode, qui perd ses propriétés émissives avec le temps. Il n'en est pas de même avec les transistors, dont on est en droit d'attendre une existence presque éternelle. De plus, les conditions qui règnent dans un matériel à transistors font que non seulement ceux-ci, mais la plupart des autres composants tombent moins souvent en panne : les puissances dissipées dans les résistances sont plus faibles, les résistances de très hautes valeurs, fragiles, n'existent plus, les capacités ont à supporter des tensions beaucoup plus petites, la température est plus basse...

les stimulateurs de cœur

Donnons encore un exemple pour illustrer les extraordinaires possibilités du transistor. Certaines personnes, atteintes d'une grave maladie de cœur, vivent avec un petit appareil implanté sous la peau par intervention chirurgicale. Cet appareil, fait avec des transistors alimentés par une pile, excite le cœur avec des impulsions électriques, rendant ainsi son fonctionnement régulier, alors que sans lui le cœur risquerait de s'arrêter et de provoquer une syncope. Ces malades, autrefois condamnés à vivre à l'hôpital avec une infirmière en alerte permanente à leur chevet, et, finalement, à mourir à plus ou moins brève échéance, ont pu reprendre une vie normale...

Lorsque le transistor est apparu, on pensait généralement qu'il viendrait compléter les tubes, qu'on l'utiliserait dans les cas où ceux-ci seraient inadaptés. Aujourd'hui, les rôles sont inversés : chaque fois qu'on a le choix entre un transistor et un tube, on pense d'abord au transistor, réservant pour le tube les emplois où le transistor ne peut convenir, emplois qui tendent à diminuer chaque jour un peu plus, car le champ d'action des semi-conducteurs s'étend sans cesse...

pour les hyperfréquences : des tubes, mais sur des principes différents

Nous avons vu tout l'intérêt qu'il y a, dans le domaine des télécommunications, à pouvoir utiliser les très hautes fréquences. Cela implique de disposer, dans le poste émetteur, d'un oscillateur capable de produire des courants sinusoïdaux aux fréquences en question, et, dans le poste récepteur, d'un amplificateur apte à amplifier les microcourants de même fréquence qui apparaissent dans l'antenne.

Or, les tubes classiques, comme les transistors, cessent de fonctionner correctement au-delà de certaines fréquences.

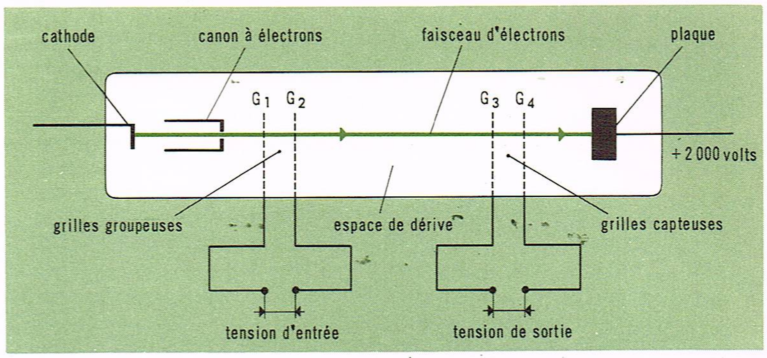
La cause principale en est, dans le cas des transistors, la présence des « charges de diffusion » dans la base (qu'on réduit en en diminuant la largeur). Dans le cas des tubes, c'est le temps que mettent les électrons à parcourir l'espace cathode-anode, ou « temps de transit » : on conçoit, en effet, qu'il faille que ce temps soit nettement inférieur à la période des oscillations à produire ou du signal à amplifier. Pour diminuer le temps de transit, on peut augmenter la vitesse des électrons en jouant sur la tension accélératrice de l'anode, ou bien diminuer les distances inter-électrodes. On arrive ainsi, avec les tubes classiques, jusqu'à environ 500 MHz. Certains tubes, où les électrodes sont disposées de façon spéciale (lampes « phare »), atteignent même plusieurs milliers de mégahertz.

fonctionner à haute puissance et à grande fréquence

Cependant, pour réaliser l'oscillateur du poste émetteur, il est nécessaire de disposer d'un élément actif qui non seulement fonctionne à la fréquence désirée, mais puisse également fournir une puissance importante à cette fréquence, ce qui est indispensable pour que l'émetteur ait une portée correcte. Or, c'est précisément la simultanéité de ces deux qualités qu'il est très difficile d'obtenir avec les composants dont nous avons

parlé jusqu'à présent. Les puissances disponibles sur les transistors, et particulièrement sur les transistors pour hautes fréquences, sont bien trop faibles (quelques centaines de milliwatts) pour qu'il soit question de les utiliser ici. Les tubes autorisent des puissances beaucoup plus grandes, mais qui décroissent lorsqu'on cherche à augmenter la fréquence. En effet, dans les tubes de forte puissance, les électrodes ont nécessairement des surfaces importantes, car la cathode doit pouvoir émettre un large flux d'électrons, et l'anode dissiper la chaleur produite par le choc de ces électrons. Si, pour diminuer le temps de transit, on rapproche les électrodes, on forme des condensateurs qui, comme le temps de transit, limitent la fréquence maximale : on ne peut donc aller très loin dans cette voie. Il faut faire appel à des principes différents. Les éléments qui permettent d'obtenir à la fois puissance et fréquence sont encore des tubes à vide, mais dans lesquels, au lieu de chercher à éviter le temps de transit, on l'utilise pour produire les oscillations. Les plus importants de ces tubes sont les klystrons, qui peuvent être soit amplificateurs, soit oscillateurs, les magnétrons, oscillateurs seulement, et les tubes à ondes progressives, amplificateurs pouvant fonctionner dans une large bande de fréquences.

La place nous manque malheureusement pour parler en détail de tous ces tubes. Aussi ne décrirons-nous que le klystron, en mettant en relief à son sujet quelques particularités de la technique des hyperfréquences.



Principe du klystron.

Du canon à électrons sort un faisceau parallèle et homocinétique. La tension alternative d'entrée est appliquée entre des grilles groupeuses G_1 G_2 : elle freine ou accélère les électrons, suivant l'instant de leur passage. Lorsque ceux-ci ont effectué un certain trajet dans l'espace de dérive, où il n'y a aucun champ, les électrons les plus rapides ont rattrapé les plus lents, des paquets se forment dans le faisceau, à la même fréquence que celle du signal d'entrée et à une distance déterminée des grilles groupeuses. On y place les grilles capteuses, reliées à un circuit extérieur, aux bornes desquelles les paquets d'électrons font apparaître un signal amplifié par rapport à celui d'entrée.

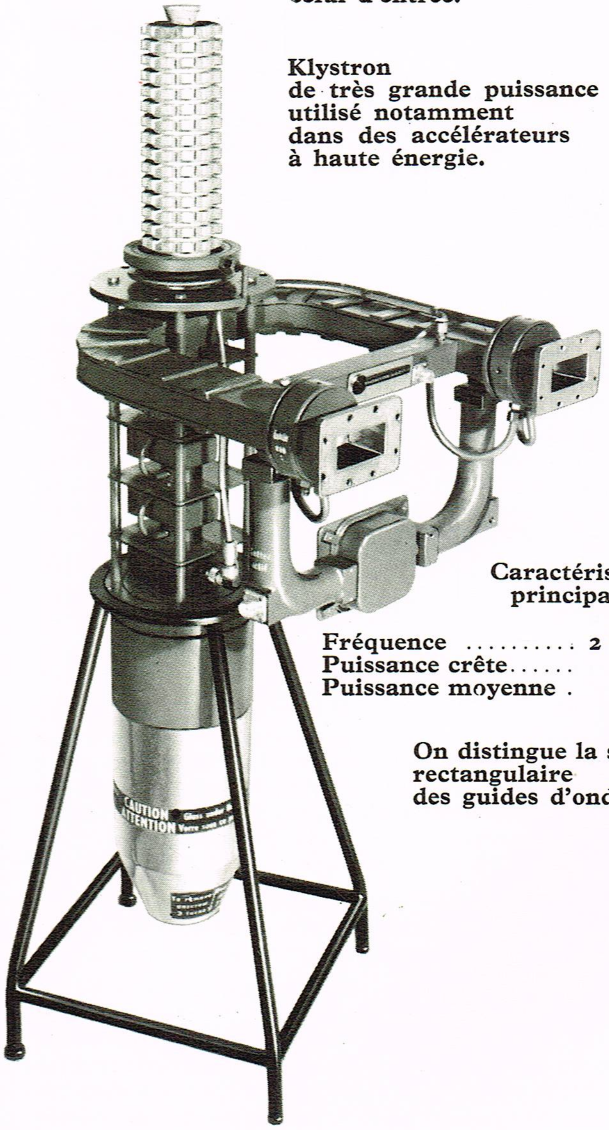
le klystron, tube hyperfréquences à modulation de vitesse

Le principe du klystron est le suivant. On réalise d'abord un faisceau d'électrons animés d'une même vitesse et dont les trajectoires sont parallèles à l'axe du tube, à l'aide d'un « canon à électrons », ensemble comprenant entre autres une cathode thermo-ionique et une ou plusieurs électrodes jouant le rôle de grilles. Les électrons passent ensuite à travers les deux grilles de commande G1 et G2 (grilles groupeuses), se propagent ensuite dans un espace vide (espace de dérive), puis traversent les deux grilles G3 et G4 (grilles capteuses) avant de se faire recueillir par la plaque, portée à un très haut potentiel (plusieurs milliers de volts par rapport à la cathode du canon à électrons). Supposons que nous désirions amplifier une tension alternative de haute fréquence. Cette tension est appliquée entre les deux grilles groupeuses, de sorte qu'il apparaît dans l'espace compris entre G1 et G2 un champ électrique alternatif de même fréquence. Il en résulte que, dans cet espace, les électrons seront soumis à une force parallèle à leur trajet dirigée tantôt dans un sens, tantôt dans l'autre : pendant la moitié de la période du signal appliqué, ils seront accélérés, et pendant l'autre moitié, ralentis. Le faisceau d'électrons qui sort de G2 a toujours une densité uniforme, car l'espace entre G1 et G2 est très petit, mais la vitesse des électrons dépend du moment où ils sont passés entre les deux grilles : le faisceau a été « modulé en vitesse » par la tension appliquée.

trois électrons qui font la course

Le problème consiste à transformer cette modulation de vitesse en modulation d'intensité, de densité du faisceau. On parvient à ce résultat en laissant simplement les électrons se propager dans l'espace de dérive, où il n'y a aucun champ, de sorte qu'ils y conservent la vitesse qu'ils avaient en sortant de G2. Essayons d'expliquer ce qui se passe, en imaginant trois électrons arrivés à des instants différents entre les grilles groupeuses : supposons que le premier ait été retardé, que le second ait gardé sa vitesse initiale (il est passé juste quand G1 et G2

Klystron de très grande puissance utilisé notamment dans des accélérateurs à haute énergie.



Caractéristiques principales :

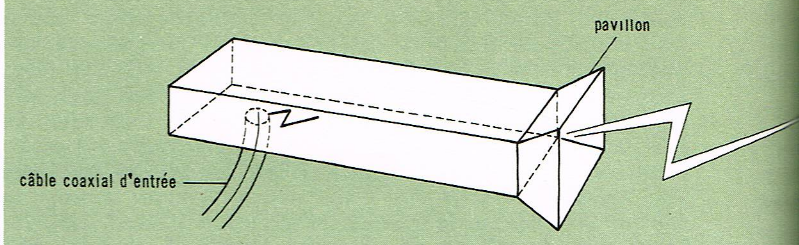
Fréquence .... 2 856 MHz
Puissance crête...... 20 MW min.
Puissance moyenne . 20 kW min.

On distingue la section rectangulaire des guides d'ondes.

Phot. Thomson-Houston.

étaient au même potentiel), et que le troisième ait été accéléré. Après une certaine distance dans l'espace de dérive, le premier électron, le plus lent, aura été rejoint par le deuxième, et également par le troisième, du moins si ces trois électrons étaient initialement séparés par des intervalles de temps égaux. On comprend ainsi comment les électrons peuvent se regrouper en paquets à une distance bien déterminée des premières grilles : il y a un paquet par période de la tension appliquée, et le reste du temps le faisceau est pauvre en électrons. C'est à cet endroit privilégié qu'on place les grilles capteuses G3 et G4, qui prélèvent de l'énergie au faisceau : en disposant entre elles un circuit extérieur convenable, les paquets d'électrons y créent des oscillations de même fréquence que celles qui sont introduites entre les grilles groupeuses. Mais, alors que la puissance nécessaire pour réaliser la modulation de vitesse est très faible, il est possible d'obtenir entre les grilles capteuses une puissance beaucoup plus grande, qui dépend surtout de l'intensité du faisceau à la sortie du canon à électrons : le klystron constitue donc un amplificateur.

En pratique, les circuits d'entrée et de sortie ne sont pas les simples grilles qui sont représentées sur le schéma, mais des cavités résonnantes. Les grilles ne permettraient pas, en effet, le fonctionnement du klystron aux très hautes fréquences; nous allons voir pourquoi.



Guide d'ondes ouvert formant l'antenne directive d'un poste émetteur à très haute fréquence.

Le signal hyperfréquence est amené par câble coaxial sur une petite antenne qui le transforme en onde électromagnétique. Celle-ci se propage dans le guide sans affaiblissement sensible en se réfléchissant sur ses parois jusqu'au pavillon, où elle est rayonnée dans l'espace.

aux hyperfréquences, pas de fil métallique

Aux basses fréquences, on transporte le courant électrique par du fil métallique, dont on considère que la résistance est nulle, pourvu que son diamètre soit suffisant. Il n'en est plus de même lorsqu'on a à transporter un courant à très haute fréquence : un simple fil métallique possède toujours une self-inductance, et présente une capacité, avec les autres fils du montage, dont les effets — insensibles lorsque la fréquence est faible — deviennent rapidement néfastes lorsqu'elle augmente, au point de rendre impossible la réalisation de certains circuits. Un oscillateur, par exemple, est souvent constitué par une self en parallèle avec une capacité, avec un composant actif pour entretenir les oscillations qui, sans lui, s'amortiraient. Si l'on veut obtenir ainsi de très hautes fréquences, on est conduit à choisir de très faibles valeurs pour la self L et la capacité C (la fréquence des oscillations étant égale à $1/2\pi\sqrt{LC}$). D'après ce que nous venons de voir, cela n'est pas possible indéfiniment, et, pratiquement, il est difficile d'aller de cette manière au-delà de 200 MHz.

De plus, un fil parcouru par un courant variable à haute fréquence émet comme une antenne des ondes électromagnétiques, d'où une perte d'énergie par rayonnement conduisant à un affaiblissement du signal, et peut également capter les ondes émises par les fils voisins, qui ne sont donc plus isolés les uns des autres.

L'emploi du « câble coaxial » permet de surmonter en partie ces obstacles. Le fil métallique est ici placé dans une gaine cylindrique, constituée en partie par un blindage métallique qui sert de retour pour le signal. Moyennant certaines conditions, le câble se présente comme une résistance pure, indépendante de la fréquence, et les pertes par rayonnement sont supprimées.

Néanmoins, on ne peut dépasser 1 500 MHz avec des câbles coaxiaux d'une certaine longueur.

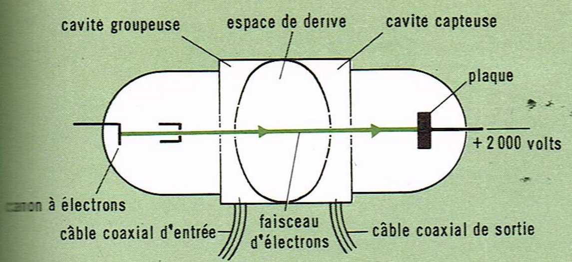
Au-delà, il faut encore faire appel à un autre principe. On ne transmet plus l'énergie électrique sous forme d'un flux d'électrons se propageant dans un métal, mais sous forme d'ondes électromagnétiques se propageant dans l'air ou dans le vide. Ce qui remplace le fil électrique est une sorte de tuyau métallique appelé « guide d'onde », dont la section est souvent rectangulaire. L'énergie électrique y est amenée par une petite antenne située à l'une de ses extrémités, reliée au générateur par un câble coaxial aussi court que possible. L'antenne rayonne des ondes électromagnétiques, mais ici elles restent confinées à l'intérieur du guide, et, si la section de celui-ci respecte certaines conditions, elles s'y propagent en se réfléchissant sur les parois métalliques avec un affaiblissement minime. On pourra recueillir l'énergie en plaçant à l'autre extrémité du guide une antenne semblable à la première, reliée au circuit d'utilisation par câble coaxial. S'il s'agit de véhiculer les oscillations fournies par un poste émetteur, l'extrémité du guide sera simplement ouverte, et constituera l'antenne émettrice.

les cavités résonnantes, organes d'entrée et de sortie du klystron

Une cavité résonnante est constituée par une enceinte métallique complètement fermée. Les ondes électromagnétiques se réfléchissent sur les parois, et pour certaines fréquences, dépendant des dimensions et de la forme de l'enceinte, il se produit un système d'ondes stationnaires, analogues à celles des cordes vibrantes, pour prendre un exemple à une dimension. Si la cavité est « excitée » à une de ses fréquences de résonance, les oscillations électromagnétiques se maintiendront en ne s'amortissant que faiblement. Les cavités résonnantes remplaceront donc, pour la production et l'amplification d'oscillations, les circuits à self-inductance et capacité utilisés en basse et moyenne fréquence.

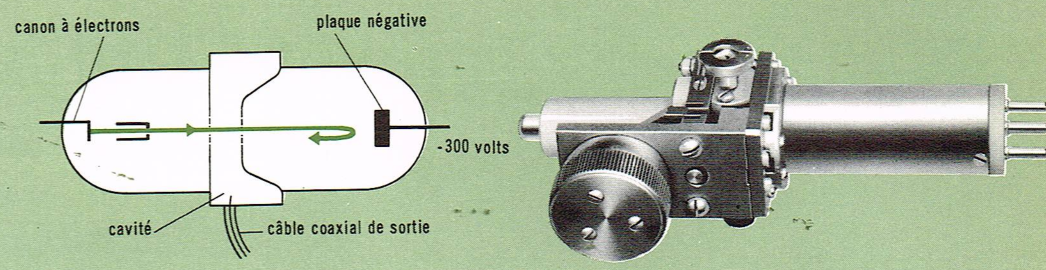
Revenons au klystron. Les deux systèmes de grilles sont remplacés par des cavités, dans lesquelles on a ménagé des orifices pour qu'elles puissent être traversées par le faisceau d'électrons. La cavité groupeuse est excitée par le signal d'entrée, à amplifier, qui est amené par câble coaxial sur une petite antenne, comme pour un guide d'onde. Le faisceau la traverse à un endroit où le champ électrique lui est parallèle et a une amplitude maximale, de façon à pouvoir moduler la vitesse des électrons. Quant à la cavité capteuse, elle entre en résonance sous l'effet des paquets d'électrons qui la traversent. Le signal de sortie en est extrait par une petite antenne reliée par un câble coaxial au circuit d'utilisation.

On peut faire du klystron un oscillateur en injectant dans la cavité groupeuse le signal de sortie. Il suffit pour cela de relier



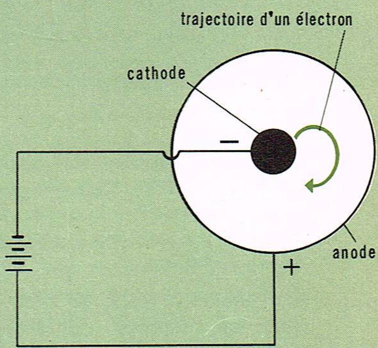
Klystron à deux cavités résonnantes.

Les grilles sont remplacées par des cavités résonnantes, dans lesquelles de petits orifices laissent passer le faisceau d'électrons. Le klystron représenté ici est un amplificateur, mais sa fréquence de travail est déterminée une fois pour toutes par les dimensions des cavités.



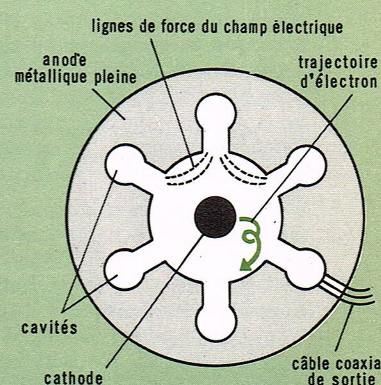
Klystron « Reflex ».

Dans ce type de klystron, uniquement oscillateur, les deux cavités sont confondues en une seule, que le faisceau d'électrons traverse deux fois.



Coupe d'un magnétron à anode pleine.

Le champ magnétique, perpendiculaire au champ de la figure, courbe le trajet des électrons qui quittent la cathode par effet thermo-ionique. Suivant qu'il est plus ou moins intense, les électrodes atteignent ou non l'anode cylindrique : le courant passe ou ne passe pas.



Magnétron à cavités.

Lorsque les cavités oscillent, chacune d'elles est déphasée d'une certaine quantité par rapport à la précédente (en général de 180°). On montre que les électrons, dont la plupart ont une trajectoire analogue à celle de la figure, entretiennent ces oscillations.

les câbles coaxiaux d'entrée et de sortie, et de faire en sorte, notamment en agissant sur la tension de la plaque, que le signal injecté respecte les conditions de phase qui permettront aux oscillations de s'entretenir d'elles-mêmes.

Un type de klystron fréquemment employé comme oscillateur est le « klystron reflex » : là, il n'y a plus qu'une seule cavité, qui joue à la fois le rôle de groupeuse et de capteuse. Les électrons la traversent une première fois, puis sont renvoyés vers la cathode par une plaque portée à un potentiel négatif qui les freine. Ils traversent ainsi une deuxième fois la cavité, mais dans le sens opposé, et s'ils le font à l'instant approprié, les oscillations s'entretiennent.

quelques remarques pour terminer

Faisons à propos du klystron quelques observations, valables pour la plupart des tubes hyperfréquences.

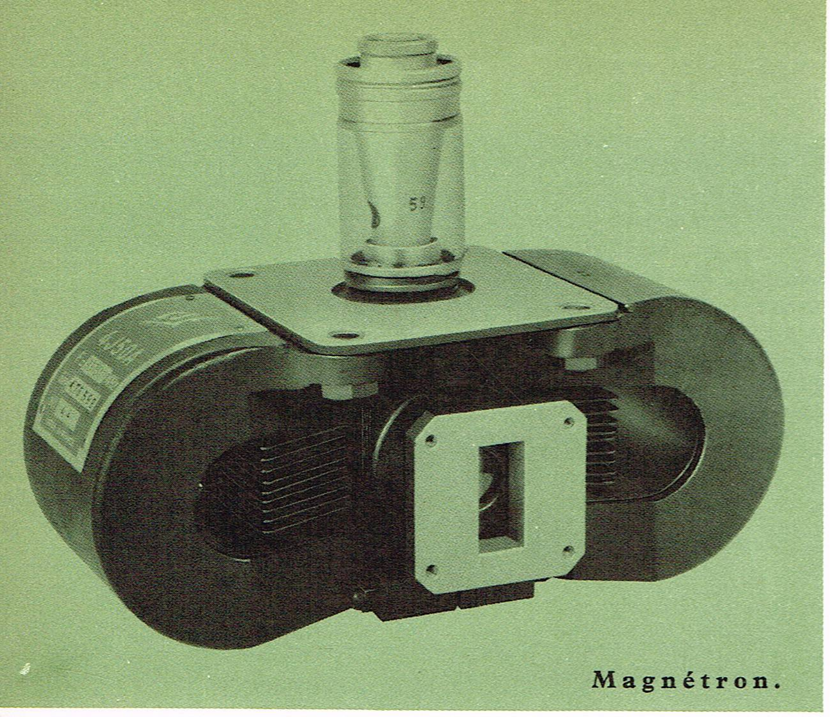
Lorsqu'on veut réaliser un oscillateur à basse fréquence, on part d'un composant actif, entre les électrodes duquel on branche le « circuit extérieur » (par exemple, self et capacité) qui définit la fréquence des oscillations, laquelle ne dépend pas du composant : si l'on veut une autre fréquence, il suffit de modifier les valeurs des éléments passifs du circuit extérieur. Il n'en est pas de même pour le klystron. D'abord, le circuit extérieur, constitué par les cavités résonnantes, n'est pas séparable du klystron lui-même. Ensuite, du fait qu'il est nécessaire d'exciter les cavités à l'une de leurs fréquences de résonance, ou du moins à une fréquence très voisine, la fréquence des oscillations est pratiquement déterminée une fois pour toutes : on arrive à la faire varier, notamment en agissant sur la tension de la plaque, mais d'une quantité toujours faible : par exemple, de 10 MHz pour un klystron fonctionnant à 3 000 MHz. **On voit que l'utilisation d'un klystron est loin d'être aussi commode que celle des tubes classiques ou des transistors : ce qu'on a gagné en fréquence et en puissance, on l'a perdu en souplesse d'emploi.**

le magnétron à cavités, oscillateur hyperfréquences de forte puissance

Le magnétron qui nous intéresse diffère fort, par sa forme et son mode de fonctionnement, du premier tube de même nom, apparu en 1921. Ce dernier est essentiellement une diode à vide placée dans un champ magnétique, le passage des électrons de la cathode à l'anode étant contrôlé par l'intensité du champ.

Le tube est cylindrique, l'anode en forme la surface et la cathode en occupe l'axe. Le champ magnétique, appliqué par exemple par un aimant permanent dans l'entrefer duquel se trouve placé le tube, est dirigé parallèlement à l'axe du cylindre. Lorsqu'un électron quitte la cathode par effet thermo-ionique, il a tendance à se diriger vers l'anode, dont le potentiel positif l'attire. Mais, dans l'espace qui sépare les deux électrodes, il est soumis à l'action du champ magnétique, qui incurve sa trajectoire. Si le champ a une valeur élevée, la courbure en est telle que l'électron retombe sur la cathode, et aucun courant ne passe dans le circuit extérieur. Si, au contraire, le

Phot. Philips.



Magnétron.

champ est faible, les électrons parviennent malgré celui-ci jusqu'à l'anode, et le magnétron fonctionne comme une diode à vide ordinaire. Il existe donc une valeur critique du champ magnétique, de part et d'autre de laquelle le courant passe ou ne passe pas.

On pourrait, avec ce type de tube, réaliser un amplificateur. Mais on s'aperçut rapidement qu'il était susceptible, à cause de la forme des trajectoires électroniques, de donner naissance à des oscillations à haute fréquence, et c'est en tant qu'oscillateur que le magnétron s'est développé et a connu le succès. Après diverses étapes intermédiaires, on est arrivé au magnétron à cavités multiples, inventé au début de la dernière guerre pour les besoins du radar, et qui est sa forme actuelle. Le fonctionnement du magnétron à cavités n'est pas simple, et tous les points n'en sont pas encore parfaitement connus.

L'anode est ici constituée par une masse métallique dans laquelle sont creusées un certain nombre de cavités identiques, communiquant avec l'espace vide cathode-anode par des fenêtres. Les oscillations sont prélevées par une petite antenne placée à un endroit déterminé d'une des cavités.

des électrons aux trajets curieux

Chaque cavité peut entrer en oscillations à une fréquence donnée. Mais ces oscillations électromagnétiques respectent une certaine relation de phase avec celles des cavités voisines, conséquence du couplage qui existe entre elles grâce aux fenêtres (en général, le déphasage entre les oscillations de deux cavités contiguës est de 180°). Le champ magnétique est réglé au voisinage de la valeur critique, un peu au-delà, de façon que la plupart des électrons n'atteignent pas l'anode.

Supposons que l'ensemble des cavités oscille, et examinons comment il est possible que les électrons issus de la cathode entretiennent ces oscillations : ce sera le cas s'ils fournissent de l'énergie, c'est-à-dire s'ils sont freinés par les champs électriques présents dans l'enceinte. Si, au contraire, ils sont accélérés par ces champs, ils absorberont de l'énergie et amortiront les oscillations.

Lorsqu'un électron sort de la cathode, il décrit une trajectoire qui l'amène, du fait de la valeur choisie pour le champ magnétique, au voisinage de l'anode, et presque tangentiellement à celle-ci. Là, il est soumis au champ électrique alternatif dû aux oscillations électromagnétiques des cavités. Il peut donc arriver à un instant tel qu'il soit accéléré, ou bien au contraire freiné : de toute façon, sa trajectoire sera modifiée. On montre que dans le premier cas il retombe rapidement sur la cathode ou l'anode, alors que dans le second il a tendance à rester dans l'espace vide, en tournant le long des cavités suivant une trajectoire plus ou moins compliquée; ainsi, ayant été freiné au voisinage d'une cavité, il pourra l'être de nouveau lorsqu'il se présentera devant la cavité suivante : s'il existe un déphasage de 180° entre les oscillations de deux cavités voisines, il suffira pour cela qu'il parcoure la distance séparant deux cavités en un temps égal à une demi-période des oscillations. Un même électron pourra donc céder une partie de son énergie à plusieurs cavités successives. En fin de compte, il y a entre la cathode et l'anode beaucoup plus d'électrons freinés, qui entretiennent les oscillations, que d'électrons accélérés, qui les amortissent. On conçoit ainsi que le magnétron à cavités puisse constituer un oscillateur.

Le magnétron est susceptible de fournir des puissances considérables pendant des durées faibles, par exemple 100 kW à 30 000 MHz pendant quelques microsecondes. Cependant, on ne peut pas agir, même faiblement, sur sa fréquence par des moyens électroniques, comme c'est le cas pour le klystron : la modulation de fréquence n'est donc pas possible, et son champ d'applications reste surtout le radar.

des tubes spéciaux pour la télévision

Une liaison par radio nécessite à la fois un élément qui transforme les signaux sonores en signaux électriques — c'est le micro du poste émetteur — et un élément qui, inversement, restitue les signaux sonores à partir de signaux électriques — c'est le haut-parleur du poste récepteur.

La télévision permet la transmission par voie hertzienne non pas d'ondes sonores, mais d'images optiques. Elle exige donc aussi des dispositifs qui assurent la transformation des signaux lumineux en signaux électriques, ainsi que la transformation inverse. Ces deux fonctions essentielles sont remplies par des tubes spéciaux dont le rôle équivaut à ceux du micro et du haut-parleur, mais qui sont beaucoup plus complexes.

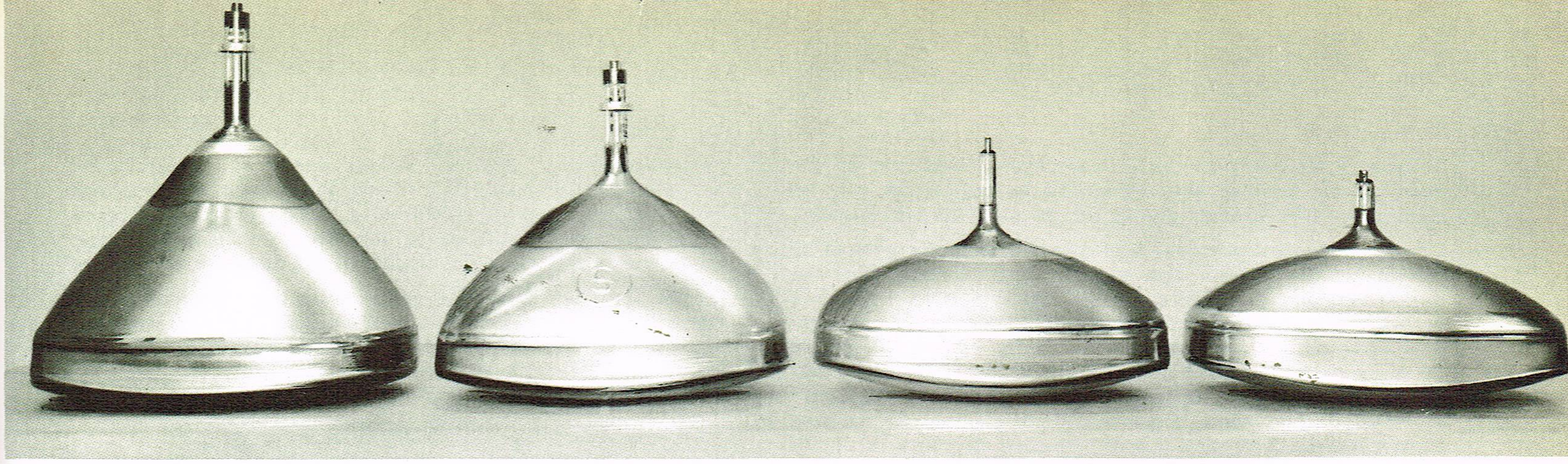
pour l'émetteur, l'effet photo-électrique

Le premier de ces tubes, celui qui a sa place dans le poste émetteur, fait appel à l'effet photo-électrique; sous l'action des radiations lumineuses, certains corps semi-conducteurs émettent des électrons d'autant plus nombreux que la lumière est plus intense (dans le cas de l'effet thermo-ionique, c'est, rappelons-le, la température qui produit l'émission des électrons). Lorsqu'on place un tel corps dans un tube à vide, dont il constitue la cathode, et que l'on porte une plaque (l'anode) à un potentiel positif, les photo-électrons sont dirigés vers celle-ci, et leur flux, c'est-à-dire le courant anode-cathode, facilement mesurable, est proportionnel à l'intensité lumineuse. On obtient ainsi une cellule photo-électrique d'un type qui a été et est encore fréquemment employé. Les tubes de prise de vues de la télévision sont évidemment infiniment plus compliqués, car ils doivent restituer sous forme électrique, non pas un point lumineux, mais une image à deux dimensions. Il en existe plusieurs sortes, fondées sur des principes voisins : l'iconoscope, l'orthicon, le vidicon...

pour le récepteur, la fluorescence

Le tube du récepteur de télévision s'appelle un « tube cathodique ». Il utilise la propriété qu'ont certaines substances, dites « fluorescentes », de devenir lumineuses lorsqu'on les bombarde par des électrons, la luminosité étant fonction de la quantité d'électrons incidents.

Phot. Bouillot (C. S. F.).



L'évolution des tubes cathodiques utilisés dans les téléviseurs a permis de construire des postes chaque fois plus plats.

Un tube cathodique comprend un écran fluorescent, et un mince pinceau d'électrons qu'on peut diriger sur celui-ci au moyen de champs électrique ou magnétique. Le point d'impact du pinceau sur l'écran définit une petite tache lumineuse, sur la position de laquelle il est donc facile d'agir. En faisant varier l'intensité du pinceau, on en règle également la luminosité.

Les tubes cathodiques ne servent pas qu'en télévision. Les oscillographes cathodiques, sur lesquels ils sont montés, permettent de visualiser des tensions et leur évolution dans le temps; ce sont des instruments de mesure fondamentaux, indispensables dans tout laboratoire d'électronique.

Avant d'examiner plus en détail les tubes de prise de vues, donnons quelques précisions sur le principe même de la télévision. Nous verrons quelles sont les différences avec la radio, et pourquoi il faut utiliser des fréquences élevées.

comment fonctionne un appareil de télévision

Transmettre par ondes hertziennes la luminosité d'un point n'est pas difficile; il suffit d'opérer exactement comme pour la radio, en remplaçant le micro par une cellule photo-électrique placée en ce point, et le haut-parleur par un tube cathodique. A l'émission, le courant de la cellule sera modulé par une fréquence radio-électrique et dirigé sur l'antenne; à la réception, le courant capté sera détecté, puis agira sur la densité du pinceau électronique du tube cathodique, faisant apparaître sur celui-ci un point dont la luminosité sera liée à celle du point origine.

Comme la télévision doit transmettre une image, c'est-à-dire une infinité de points inégalement lumineux juxtaposés, on pourrait imaginer un tube de prise de vues constitué par autant de petites cellules que de points élémentaires, occupant une surface sur laquelle, par un système optique, serait projetée la scène à téléviser; le courant de chaque cellule serait véhiculé par voie hertzienne comme précédemment, et, après détection, agirait sur la luminosité d'un point fixe de l'écran du récepteur, homologue par sa position à la cellule du tube de prise de vues. Malheureusement, il faut au moins 300 000 points pour définir une image correcte. La méthode précédente exigerait donc 300 000 fréquences différentes, et autant d'émetteurs et de récepteurs; elle est irréalisable.

grâce à la persistance rétinienne

On tourne la difficulté en profitant d'une propriété de l'œil humain : la persistance rétinienne. Au lieu de transmettre **simultanément** les luminosités des points élémentaires de l'image, on le fait **successivement,** ce qui ne nécessite plus qu'un seul canal radio-électrique. Supposons que le tube de prise de vues soit toujours formé d'un grand nombre de petites cellules photo-électriques, et qu'il soit possible de les interroger, d'en extraire l'information, en dirigeant sur elles un faisceau d'électrons. Nous verrons tout à l'heure comment se fait cette opération; pour l'instant contentons-nous d'admettre que, lorsque le faisceau d'électrons vient frapper une cellule, il apparaît quelque part un courant proportionnel à l'éclairement de celle-ci. Admettons également qu'on puisse diriger le faisceau sur chaque point de l'image, de la même façon que dans un tube cathodique. Le principe consiste à explorer les 300 000 cellules les unes après les autres et à synchroniser le mouvement du faisceau du tube récepteur avec celui du tube de prise de vues : à chaque instant apparaît sur l'écran du récepteur un point dont la luminosité est proportionnelle à celle de la cellule correspondante du tube émetteur. Lorsque l'exploration, ou « balayage », est terminé, on recommence.

mais il faut des fréquences élevées

Mais, pour que l'image apparaisse en continu sur l'écran, et non comme une succession de points, il faut que les balayages se succèdent à un rythme suffisamment rapide. Il faut d'autre part restituer la sensation de continuité du mouvement. C'est ici que la persistance rétinienne intervient : si la cadence de formation des images est assez élevée, l'œil intègre les petites différences discontinues qui existent entre elles (c'est ce qui se passe pour le cinéma, qui consiste à projeter successivement des vues fixes). Bref, tout ceci impose une fréquence minimale de balayage, qui est de l'ordre de 30 par seconde. Le faisceau d'électrons ne restera donc sur chacune des 300 000 cellules que pendant un très court instant, exactement $\frac{1}{30 \times 300\,000}$ de seconde, soit environ 0,1 microseconde. Ce qui signifie que le courant représentant la luminosité devra varier à la fréquence de 10 millions de fois par seconde (10 MHz). Or, c'est ce courant qu'il faut moduler et transporter par ondes hertziennes; c'est pourquoi la fréquence de l'onde porteuse, qui doit être nettement supérieure à la fréquence maximale du signal à transmettre, est nécessairement très élevée en télévision, près de mille fois plus grande que la fréquence minimale permettant de transmettre les signaux sonores, limités à 15 000 Hz. En France, les divers émetteurs fonctionnent autour de 200 MHz.

les tubes de prise de vues, chefs-d'œuvre de la technique

Nous décrirons ici l'iconoscope, dont les autres tubes diffèrent surtout par des perfectionnements technologiques.

Comme tous les tubes de prise de vues, l'iconoscope comporte : une surface recouverte d'une couche présentant l'effet photo-électrique; un système optique qui projette sur cette couche

Phot. Nobécourt.

la scène à téléviser; un canon à électrons fournissant le faisceau; et des bobines électromagnétiques permettant, par les champs qu'elles créent, d'agir sur la direction du faisceau.

La couche photo-électrique, faite d'un mélange d'argent et de césium, est déposée sur une mince plaque de mica, puis craquelée à chaud de façon à obtenir une « mosaïque » à très grande finesse, représentant l'ensemble des cellules élémentaires dont nous parlions tout à l'heure. L'autre côté de la plaque de mica est entièrement métallisé, et branché sur une résistance extérieure. Un dispositif optique projette l'image sur la mosaïque elle-même. Comme chaque élément en est isolé électriquement, il ne peut y avoir émission de courants photo-électriques continus, mais seulement de quantités déterminées de photo-électrons : chaque cellule de la mosaïque acquiert donc une charge positive proportionnelle à son éclairement.

Lorsque le faisceau d'électrons rencontre une cellule, il en chasse, par « émission secondaire », d'autres électrons de moindre énergie. (L'émission secondaire est un phénomène comparable à l'effet photo-électrique, les photons incidents étant remplacés par des électrons; d'ailleurs, tous les corps présentant l'effet photo-électrique peuvent donner aussi des électrons secondaires, la réciproque n'étant pas toujours vérifiée. Tout cela sera dans le second tome).

Ainsi, lorsqu'une cellule est touchée par le faisceau, il en sort un certain nombre d'électrons, et ce nombre est fonction de la charge initiale de la cellule, c'est-à-dire de son éclairement. Comme une cellule forme un petit condensateur avec la face métallisée de la plaque de mica, chaque courant instantané d'électrons secondaires est transmis à la résistance extérieure à travers ce condensateur. On recueille donc, aux bornes de la résistance, une tension proportionnelle à l'éclairement de la cellule frappée par le faisceau.

La direction du faisceau dépend de la valeur des champs électromagnétiques produits par des bobines de déviation; on sait qu'un champ magnétique soumet un courant électrique à une force perpendiculaire au courant et au champ. Ici, le courant est le faisceau d'électrons. Sous l'action de la force électromagnétique, il sera dévié dans un sens déterminé. Il y a généralement quatre bobines de déviation, produisant deux champs perpendiculaires entre eux et au faisceau. Un de ces champs correspond à la déviation horizontale, l'autre à la déviation verticale. Le balayage s'obtient en envoyant dans chaque bobine un courant qui varie dans le temps suivant des lois déterminées.

quels sont les signaux transmis par voie hertzienne ?

Diverses formes de balayage ont été expérimentées; aujourd'hui, tous les pays ont adopté le balayage par lignes.

Le faisceau d'électrons explore la mosaïque suivant des lignes horizontales parcourues à vitesse constante. Lorsqu'il est arrivé au bout d'une ligne, il revient en un temps extrêmement bref au début de la ligne inférieure. Enfin, lorsque le faisceau est arrivé à l'extrémité de la dernière ligne, il est renvoyé, toujours en un temps très court, à l'origine de la première ligne.

Les tensions de commande des bobines sont des tensions en « dents de scie ». Pendant la durée d'une dent de scie du balayage vertical, il y a autant de dents de scie de balayage horizontal qu'il y a de lignes.

synchroniser le faisceau du récepteur avec celui de l'émetteur

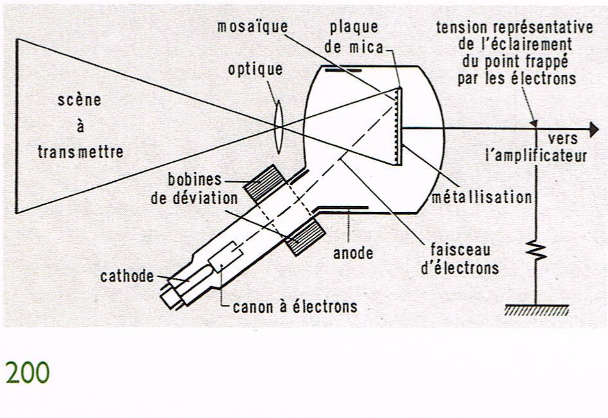
Il reste maintenant à définir comment on synchronise le mouvement du faisceau électronique du tube récepteur avec celui du tube de prise de vues.

Le faisceau d'un tube cathodique se commande de la même façon que celui de l'iconoscope, avec des bobines électromagnétiques. Le poste récepteur comporte, comme le poste émetteur, des circuits électroniques qui fabriquent des tensions en dents de scie (générateurs de balayage) permettant d'effectuer le même type d'exploration par lignes à vitesse constante. Pour assurer la synchronisation, il suffit de faire démarrer les dents de scie du récepteur en même temps que celles de l'émetteur. Pour cela, on transmet par voie hertzienne, aux générateurs de balayage du récepteur, des impulsions à la fin de chaque dent de scie de l'émetteur, impulsions qui déclenchent les dents de scie du récepteur.

En résumé, les signaux électriques qu'il faut transmettre de l'émetteur au récepteur sont de trois sortes :
— le signal de vision, fourni par le tube de prise de vues, dont la fréquence peut atteindre plus de 10 MHz par seconde;
— les impulsions de fin de ligne (impulsions de synchronisation « ligne »);
— les impulsions de fin de balayage (impulsions de synchronisation « image »).

Ces signaux étant échelonnés dans le temps (au moment où l'on envoie une impulsion de fin de ligne, il n'y a plus de signal de vision), il est facile de les « mélanger » et de les transmettre sur la même longueur d'onde. A la réception, des circuits appropriés séparent ces différents signaux et les aiguillent soit vers la commande de luminosité du tube cathodique, soit vers les générateurs de balayage.

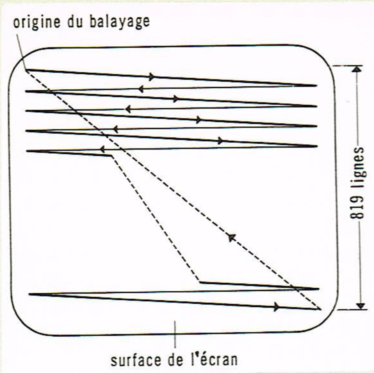
En pratique, le signal de fin de ligne est une petite impulsion rectangulaire, celui de fin de balayage une impulsion rectangulaire de durée plus grande (la différence de durée de ces impulsions suffit à les différencier à la réception). Entre deux lignes successives comme entre deux balayages, il s'écoule un certain temps pour permettre à l'impulsion de synchronisation de se former.



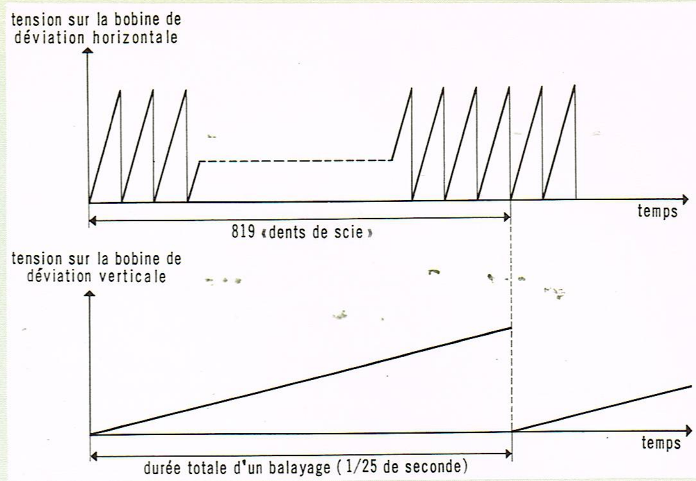
Principe de l'iconoscope.

L'image de la scène à transmettre est projetée sur la couche photo-électrique, laquelle est morcelée en une « mosaïque » d'éléments indépendants, dont chacun prend une charge électrique qui dépend de son éclairement.

Le faisceau d'électrons explore successivement tous les éléments de la mosaïque. Chaque fois qu'il en frappe un, il en chasse, par « émission secondaire », un certain nombre d'électrons, qui dépend de la charge de l'élément, c'est-à-dire de son éclairement. Ces derniers électrons donnent une impulsion de courant. qui est transmise à la face métallisée de la plaque de mica par effet condensateur. La tension qui apparaît sur cette plaque est donc significative de l'éclairement de l'élément touché par le faisceau au même instant.



Balayage par lignes et forme des tensions sur les bobines de déviation.



La durée d'exploration d'une ligne est égale à la durée totale du balayage divisée par le nombre de lignes, soit ici environ $\frac{1}{25 \times 800}$ seconde (50 microsecondes). Lorsque le faisceau est arrivé au bout d'une ligne, il revient en un temps extrêmement bref à l'origine de la ligne suivante. De même, lorsqu'il a terminé le balayage complet de l'écran (il est alors en bas, à droite), il est renvoyé, toujours très rapidement, à l'origine des balayages (en haut, à gauche).

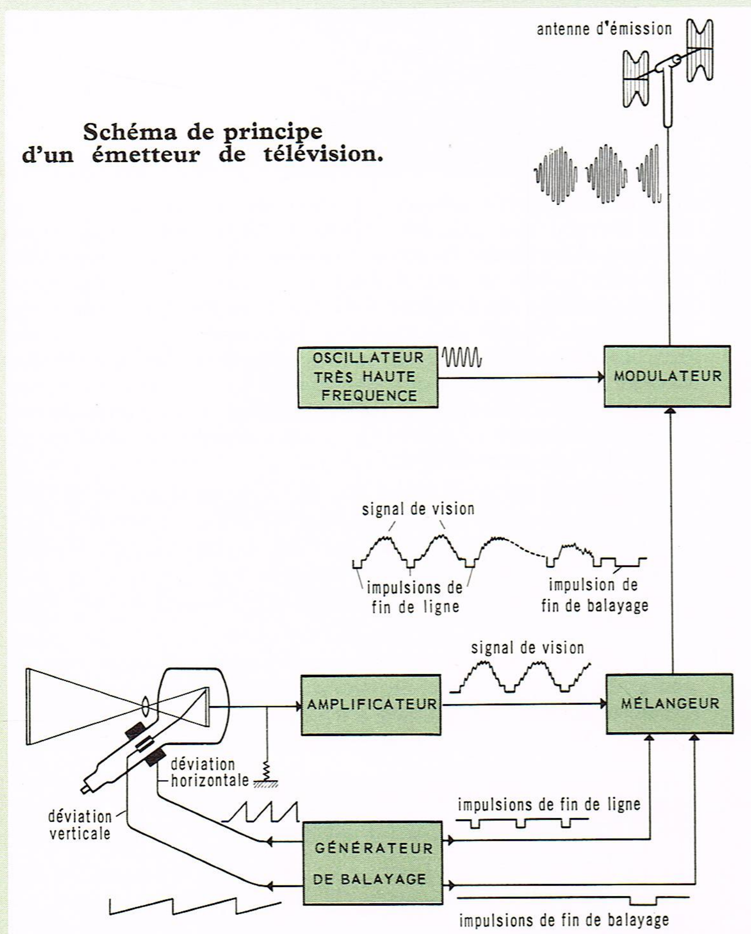


Schéma de principe d'un émetteur de télévision.

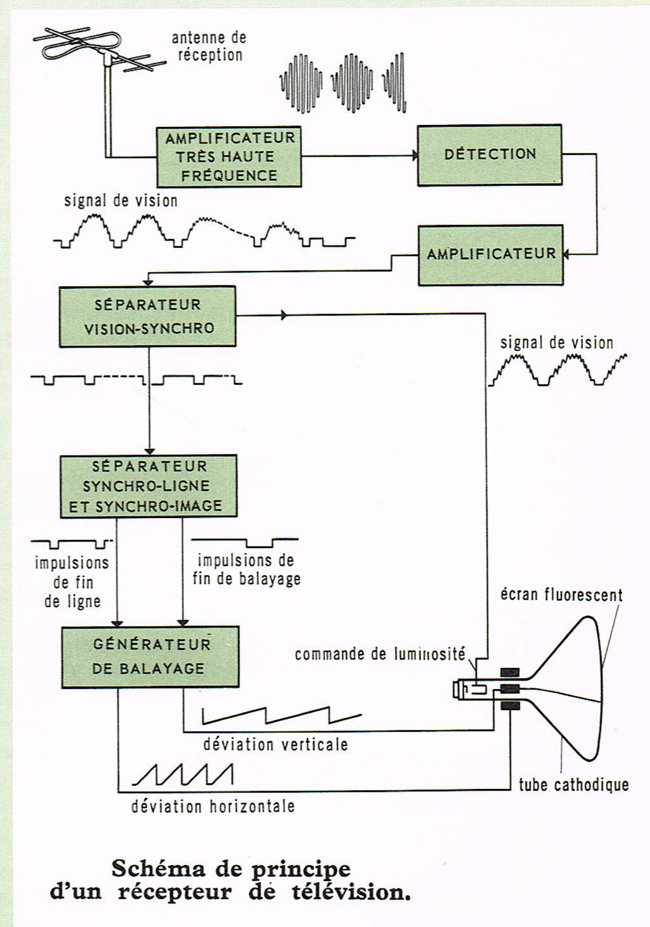


Schéma de principe d'un récepteur de télévision.

● **Emetteur de télévision. On remarquera que les impulsions de synchronisation correspondent à une absence de courant de haute fréquence dans l'antenne. Les impulsions de fin de ligne se distinguent des impulsions de fin de balayage par leur différence de durée.**

● **Récepteur de télévision. L'antenne capte le « mélange » des trois signaux : signal de vision, impulsions de fin de ligne, impulsions de fin de balayage. Après détection, des circuits appropriés séparent ces signaux : le signal de vision commande l'intensité du faisceau d'électrons, soit la luminosité du point qu'il frappe sur l'écran, les impulsions de fin de ligne déclenchent les « dents de scie » du générateur de balayage horizontal, et celles de fin de balayage déclenchent celles du balayage vertical.**

peut-on prévoir l'avenir ?

Que nous réserve l'électronique de demain? Il est bien dangereux de vouloir répondre à cette question. Tout va trop vite dans cette technique; les pronostics qu'il faudrait émettre nous feraient passer actuellement pour un tenant de la science-fiction, quitte à se révéler anachronique dans quelques années. Et comment prévoir l'apparition d'éléments nouveaux, capables, comme le transistor, de tout remettre en cause?

On peut, cependant, se hasarder, sans trop de risques, à faire quelques remarques concernant l'avenir immédiat.

les transistors, éléments actifs de plus en plus universels

Il est probable que les transistors bénéficieront de progrès technologiques permettant d'étendre leur domaine d'applications simultanément à des fréquences et à des puissances accrues.

Il est intéressant de noter comment les transistors à jonction ont conquis, petit à petit, la plupart des domaines d'applications où les tubes à vide avaient jusqu'alors permis tous les progrès de l'électronique.

D'abord limités en fréquence, les transistors ont été progressivement capables de fonctionner jusqu'à plusieurs centaines de mégahertz et dépassent largement aujourd'hui les 1 000 MHz. Les puissances possibles dans ces domaines sont en continuels progrès et atteignent quelques dizaines de watts, tandis qu'en basse fréquence plusieurs centaines de watts sont couramment obtenus. Une caractéristique importante était toutefois restée longtemps l'apanage des tubes : c'est celle de la haute impédance d'entrée. Cet avantage tombe maintenant avec les transistors à effet de champ, qui, à une impédance d'entrée aussi élevée, joignent de meilleures performances en ce qui concerne la tenue aux vibrations et la dérive à long terme.

Les tubes à vide disparaîtront-ils définitivement? Il leur reste encore un bastion difficile à conquérir par les transistors : celui des très grandes puissances en haute fréquence.

Stimulés par la concurrence du transistor, les tubes ne baissent pas pavillon. Ils deviennent plus petits, leur consommation diminue, ils s'usent moins vite. Outre leur aptitude à supporter des températures très élevées, ils possèdent un avantage indiscutable : ils sont insensibles aux radiations. A notre ère atomique et spatiale, cet aspect a son importance et leur assurera une place dans l'équipement des satellites et véhicules interplanétaires.

pour les télécommunications : lasers et masers

L'émission des ondes hertziennes reste pour l'instant l'apanage des tubes. Mais, même ici, les semi-conducteurs commencent à se manifester; avec des diodes convenablement polarisées, dont on utilise la propriété de présenter une capacité qui varie avec la tension appliquée, on arrive à faire des oscillateurs à très haute fréquence dont la puissance est suffisante pour les câbles hertziens. Comme, avec le même type de diode, on réalise aussi l'amplification haute fréquence de la partie réception, il est dès à présent possible d'équiper un câble hertzien avec des semi-conducteurs. Remarquons que, dans ce cas, la puissance exigée à l'émission est relativement faible, car la portée est pratiquement limitée à la portée optique. Il n'en est pas de même lorsqu'il s'agit d'envoyer des signaux à un satellite! Là, c'est, jusqu'à nouvel ordre, un tube hyperfréquence qu'on utilisera.

Mais la nouveauté la plus significative, dans le domaine des ondes hertziennes, est constituée par l'apparition des dispositifs à émission stimulée, appelés « masers » et « lasers », décrits dans le chapitre précédent. Les masers permettent l'émission ou l'amplification d'ondes électromagnétiques de fréquence extrêmement élevée, allant de quelques milliers de mégahertz jusqu'à l'infrarouge. Quant aux lasers, ce sont des masers travaillant avec les ondes lumineuses.

Les masers ont déjà trouvé des applications spectaculaires : ce sont eux qui se chargent de l'amplification des signaux en provenance des satellites « Telstar », « Relay », « Molnya ». Le maser est, en effet, beaucoup plus sensible que tout autre type d'amplificateur qu'on aurait pu envisager. Ainsi, lorsqu'on regarde à la télévision une émission en direct des Etats-Unis, c'est à un maser bien placé qu'on le doit.

Les lasers commencent, eux aussi, à transmettre des informations, et les procédés de modulation actuellement en cours de mise au point permettront de gagner plusieurs ordres de grandeur dans la quantité d'information émise par les faisceaux hertziens actuels ou de porter de plus en plus loin dans l'espace lors des voyages interplanétaires futurs.

une nouvelle révolution : les microcircuits

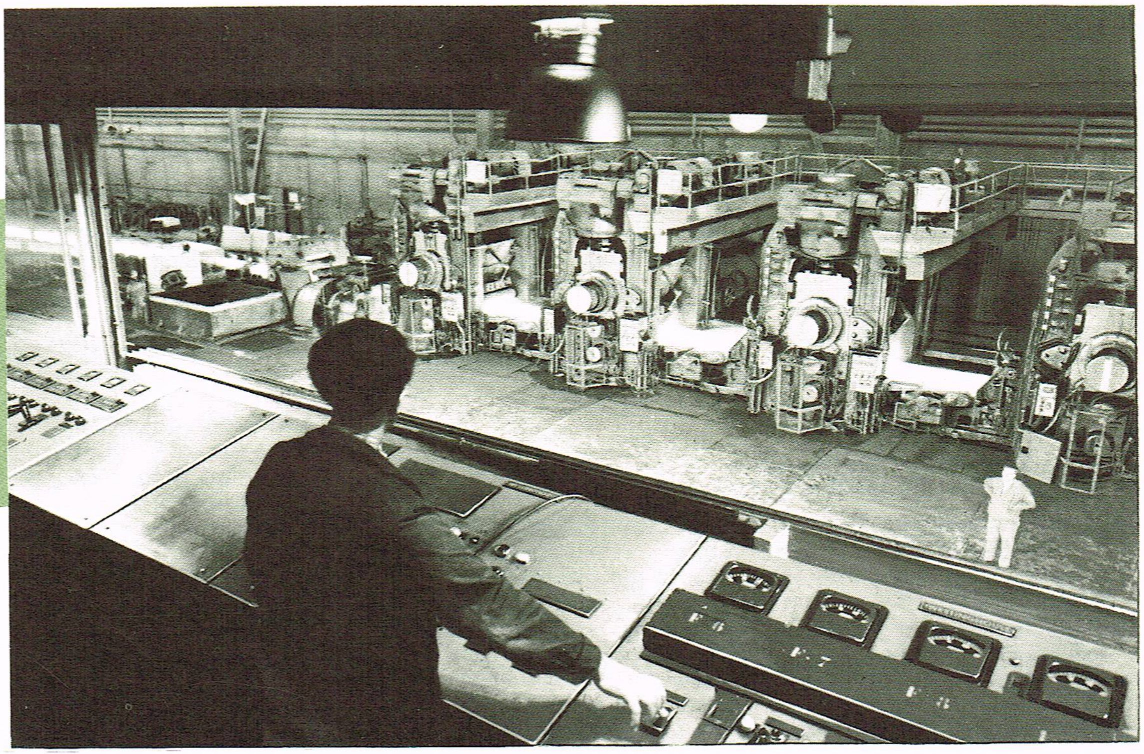
Il nous faut maintenant parler des microcircuits, dont l'avènement en tant que produits industriels de grande diffusion est en train de modifier fondamentalement les structures de l'électronique. On sait, en effet, fabriquer des circuits complets tenant dans le volume d'un boîtier de transistor. Dans ce boîtier, on peut mettre jusqu'à une centaine d'éléments divers : transistors, diodes, résistances et même capacités. De tels microcircuits ne sont nullement des curiosités de laboratoire et, pratiquement, toutes les grandes sociétés qui s'intéressaient à la fabrication des composants actifs lancent des chaînes de fabrication de microcircuits.

Il est essentiel de remarquer que les considérations de dimensions et de poids ne sont pas les seules à fixer l'intérêt sur les microcircuits. En effet, par rapport aux mêmes circuits réalisés de façon classique, la fiabilité est très augmentée et le prix de revient lui-même sera dans un avenir très proche tout à fait compétitif.

Nous allons essayer d'expliquer succinctement quelles sont les méthodes qui permettent de fabriquer ces éléments — la perfection technique atteinte vaut la peine qu'on s'y attarde un peu.

diffusion et photogravure d'un circuit intégré

La figure ci-contre représente, vus de dessus, quelques microcircuits montés sur leur socle, le capot ayant été enlevé. Le circuit lui-même, composants et connexions, tient tout entier dans le petit carré central, pastille de silicium de 1 mm de côté. Les connexions avec l'extérieur (tensions d'alimentation, signaux d'entrée et de sortie) sont assurées par l'intermédiaire des bornes qui entourent le carré de silicium : on distingue les minces fils d'or qui relient chacune d'elles avec des points déterminés du circuit. L'espèce de quadrillage qu'on aperçoit sur le circuit en constitue le « câblage » : c'est le réseau de fils métalliques qui assure les liaisons électriques entre les différents composants. Quant à ces derniers, ils sont invisibles, étant tous constitués par de minuscules volumes de silicium noyés à l'intérieur de la petite plaquette.



Grâce à l'électronique, toutes les commandes et tous les contrôles de ce laminoir sont réunis dans un poste central.

Initialement, la plaquette est un morceau homogène de silicium P dont la surface comporte une mince couche de silicium de type N, qui est lui-même recouvert d'une pellicule d'oxyde. C'est cette couche N qui constituera les collecteurs des transistors (ceux-ci sont du type NPN). Afin d'isoler les transistors entre eux, on découpe dans la couche d'oxyde, par un procédé faisant appel à la photogravure, de minces bandes sous lesquelles le silicium est mis à nu. Puis, par une diffusion à haute température, on introduit des impuretés P qui pénètrent dans le cristal par les bandes précédemment formées et rejoignent la couche inférieure du même type. On dispose ainsi de parallélépipèdes appelés « caissons d'isolement » dont les dimensions sont de quelques centièmes de millimètre. A la fin de la diffusion, on introduit de l'oxygène, et la couche d'oxyde se reforme aux endroits où elle avait été enlevée. Le deuxième cycle d'opérations consiste à diffuser de nouveau les impuretés P dans le silicium N, mais elles serviront cette fois à faire les bases des transistors et les résistances (un barreau de silicium P peut fort bien constituer une résistance). Comme précédemment, la couche d'oxyde est enlevée aux endroits prévus pour la diffusion. Elle se reforme à la fin de la diffusion, ce qui permet d'éviter toute contamination du cristal.

Enfin, les émetteurs des transistors sont obtenus au cours d'un troisième cycle par diffusion d'impuretés N. Les composants du circuit sont terminés; il ne reste plus qu'à les relier par quelque chose d'équivalent à du fil électrique. Pour cela, on dépose une mince couche d'aluminium par évaporation sous vide sur toute la surface du circuit, puis, toujours à partir de la photogravure, on n'en laisse subsister que ce qui est nécessaire à l'interconnexion.

Il va de soi que l'on ne fait pas ces éléments un à un; des centaines peuvent être réalisés simultanément sur un disque de silicium unique, et on les découpe lorsque les circuits sont terminés pour monter chacun d'eux sur un support et souder les fils de sortie. C'est justement cette possibilité d'en fabriquer un grand nombre à la fois, associée au perfectionnement constant des techniques de fabrication, qui est à l'origine de l'effondrement des prix que l'on connaît actuellement.

Le type de circuit que nous venons de décrire s'appelle un « circuit intégré ». Mais il existe d'autres méthodes pour arriver au microcircuit. Une des plus prometteuses consiste à déposer, sur une plaquette isolante, les microcomposants par évaporation sous vide. Actuellement, on réalise industriellement des résistances (très mince film métallique) et des condensateurs; des diodes et des transistors aux performances étonnantes sont fabriqués en laboratoire.

L'avènement des circuits intégrés pose également des problèmes aux utilisateurs. Pour bénéficier pleinement de la miniaturisation, ceux-ci sont obligés de revoir entièrement leur technologie de câblage. La figure de la page 204 montre quelques boîtiers



Elément micrologique à 8 transistors et 12 résistances. La fourmi donne l'échelle.

Phot. Lacheroy et Fairchild.

de circuits intégrés et un circuit imprimé de 10 × 20 cm comportant 45 circuits intégrés, dont certains ont leur boîtier en plastique. Les liaisons électriques se font sur les deux faces. Les fonctions disponibles dans un boîtier deviennent de plus en plus complexes, de même les interconnexions entre boîtiers. Aussi fabrique-t-on des circuits imprimés comportant plusieurs couches d'interconnexions. Des circuits comportant jusqu'à 10 ou 11 couches ont pu être obtenus.

La complexité d'un circuit intégré a-t-elle une limite? Certainement, mais aujourd'hui il est impossible de la chiffrer. Le circuit que nous avons décrit plus haut pourrait contenir plusieurs centaines de composants. On envisage, dès maintenant, la réalisation de circuits intégrés comportant plusieurs milliers de transistors du type « MOS », dont la phase finale d'interconnexion serait dirigée par un calculateur.

une véritable mutation

La miniaturisation et l'économie permises par l'emploi de ces « circuits intégrés » permettraient de loger, dans un appareil dont les dimensions sont bien souvent fixées par des impératifs humains tels que : dimensions de haut-parleur dans les récepteurs ou des tubes aménagés dans les téléviseurs, des circuits d'une complexité sans cesse croissante. Les informations, bien souvent encore transmises et traitées d'une manière analogique dans les appareils de grande diffusion, le seront de plus en plus d'une manière numérique. On assistera donc à la possibilité de profiter de toutes les techniques du traitement numérique de l'information actuellement réservées aux machines électroniques, et cela avec des prix de revient qui, il y a seulement cinq ans, auraient stupéfié les plus audacieux.

D'ailleurs, cette quantification de l'information permettra l'interconnexion de plus en plus intime de tous les appareils électroniques utilisés, non seulement dans un dessein professionnel, mais aussi dans la vie courante. Il suffit de citer l'expérience actuellement en cours à Phœnix, où la General Electric a doté certains domiciles de ses ingénieurs de calculateurs ménagers permettant aux épouses de faire automatiquement leur compte et de recevoir des conseils culinaires!

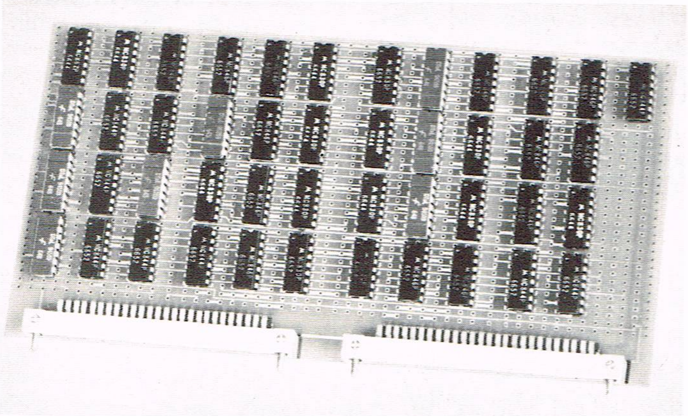
Munie de ces armes nouvelles, jusqu'où s'étendra l'électronique? A vrai dire, on ne voit guère de limite à son champ d'action. On imagine mal qu'une branche de la science ou de la technique puisse lui rester étrangère. Aux Etats-Unis, l'électronique occupait, au début de la Seconde Guerre mondiale, le quarantième rang parmi les industries productrices; aujourd'hui, elle est dans les cinq premières, et certains pensent que d'ici à dix ans elle aura dépassé toutes les autres.

une expansion que rien ne pourra limiter

Il n'est pas difficile de prévoir que partout où l'électronique est aujourd'hui présente, son rôle continuera à progresser, en particulier dans ses quatre principaux secteurs d'activité actuels, qui sont, par ordre d'importance : la défense nationale, le matériel « grand public » (radio, télévision, magnétophones), les calculateurs arithmétiques et l'électronique spatiale. Le domaine qu'elle est en train de coloniser actuellement est celui de l'électronique industrielle, en entendant par là les dispositifs de mesure, d'alarme, de régulation et de contrôle, donc tout ce qu'englobe le mot « automation ».

L'introduction de l'électronique dans les applications industrielles a longtemps été freinée par la traditionnelle méfiance des utilisateurs. En effet, quand on a sous les yeux un appareillage électromécanique constitué de moteurs, engrenages, relais, voltmètres, on croit voir ce qui se passe, on comprend à peu près. Devant les appareils électroniques, un non-initié ne voit rien, ne comprend rien, il est obligé de faire confiance à son fournisseur. Avec l'avènement des calculateurs numériques, l'abstraction est encore plus grande, puisqu'il n'y aura bientôt plus à voir que des trous sur des cartes perforées ou tout autre support similaire.

Cependant, un mouvement irréversible est amorcé. En effet, il y a quelques années un industriel lançait un petit appareil de contrôle sur le marché en fondant sa publicité sur le fait qu'il n'y avait dans cet appareil rien d'électronique. Il suffit de regarder les annonces actuelles pour se persuader de l'évolution de ces dernières années. De très nombreuses firmes sont effectivement en train de rendre électronique, en utilisant des techniques numériques et des circuits intégrés, l'appareil de mesure le plus universellement répandu : le contrôleur universel. Ainsi, la nouvelle électronique est en train de s'attaquer à l'« ancêtre » qui pourtant lui a donné le jour et l'a soutenue dans son adolescence : le galvanomètre.



Plaquette à circuits intégrés.

Cette plaquette de 10×20 centimètres contient plus de 250 transistors et un millier de composants divers (résistances, condensateurs).

Phot. X.

COMMENT SE TRANSMET ET SE DÉGRADE L'INFORMATION

par
Jacques CLAVIER

Nul doute que l'électronique ne joue un rôle immense dans notre vie. Il n'est, pour en prendre pleine conscience, que de revivre le « plaisant » tableau que l'on a pu faire d'un monde où les électrons seraient en grève : privés d'appareils tels que transistors, télérécepteurs, électrophones, abandonnés des instruments de contrôle de la circulation urbaine, des réseaux ferroviaires et aériens, incapables de se déplacer autrement qu'à pied, contraints à déserter les usines trop automatisées, n'ayant comme consolation que celle de l'impossibilité de la guerre thermonucléaire et que le nécessaire retour à la paix et au calme des champs, les hommes seraient ramenés, à la suite d'une récession géante, à l'âge barbare.

Mais, finalement, les composants électroniques ne servent qu'à transmettre de l'information. Et ce, beaucoup plus rapidement et plus sûrement que nos organes des sens, encore qu'il faille bien que toutes les connaissances que nous possédons passent une fois ou l'autre par eux.

Qui ne sait, ou ne croit savoir, ce que signifie « information ». Et ce mot fait partie de notre vocabulaire le plus courant.

Mais un concept n'a de sens et de valeur scientifiques que lorsqu'on en a précisé exactement la signification. Et la définition initiale, « renseignement, nouvelle donnés par un journal, par la radio ou la télévision », n'est ni assez générale ni assez claire pour être ici de quelque valeur. Il a fallu attendre 1948 pour que Shannon, radio-électricien spécialiste de problèmes de télécommunication, fondât scientifiquement la théorie de l'information, donnant ainsi une assise solide à beaucoup de théories jusqu'alors dispersées; une discipline d'origine très technique, puisqu'il s'agissait de coder certains messages de manière optimalement efficace, s'est révélée beaucoup plus fondamentale et profonde qu'on eût pu le supposer à première vue. De fait, il n'existe pas de science pouvant y échapper, que ce soit la thermodynamique, la physique moderne, l'art de l'expérimentation et de l'observation — c'est-à-dire finalement toute la physique —, la recherche de l'action optimale et des règles les plus efficaces de conduite (ou recherche opérationnelle); les lois même de la vie y font de plus en plus appel et les biologistes centrent de plus en plus leurs recherches sur l'information génétique.

Il s'agit donc d'une science clé de la physique moderne, et c'est la raison pour laquelle elle sera très longuement reprise dans le volume consacré aux fondements de la science; savoir d'une part que toute information est limitée, et que toute connaissance ne peut que se dégrader au cours de son nécessaire transport, que d'autre part l'homme ne peut vivre qu'en acquérant de l'information — donc du passé au futur —, est nécessaire à la compréhension de toute la physique probabiliste moderne; c'est bien parce qu'elle a ignoré ou négligé ces « petites » vérités que la physique du XIXe siècle a dû finalement céder la place à une science plus riche et plus complexe, et aussi plus difficile et moins intuitive.

En renvoyant donc le lecteur au chapitre « Information et mesure de la connaissance » (tome II) pour une étude plus approfondie, nous nous bornerons ici à une rapide synthèse des notions essentielles destinée à ceux qui veulent comprendre comment et combien électronique et information sont liées, et comment fonctionnent les machines à traitement automatique de l'information.

ce qu'est l'information

Chaque message que nous recevons, qu'il s'agisse du télégramme nous apprenant la naissance d'un cousin, du journal nous annonçant une révolution en Amérique du Sud ou un discours du chef de l'Etat, de la mercuriale du jour ou des cours de la Bourse, d'une image télévisée ou de la vue d'un paysage, nous enrichit plus ou moins en modifiant nos possibilités d'action ou en en créant de nouvelles; nous sommes incités, par exemple, à faire tel voyage ou à y renoncer, à fixer nos prochains rendez-vous en fonction des nécessités et des risques de certains voyages, à vendre ou à acheter des actions, à désirer passer nos prochaines vacances dans tel site qui paraît enchanteur, etc.

Il importe aux scientifiques de rejeter tout aspect subjectif et de ne retenir que le caractère mesurable de l'information. Que nous ayons acheté un billet ou non n'influe pas sur la quantité d'information contenue dans la phrase « le numéro 371 423 a gagné au dernier tirage de la Loterie nationale »; ce qui compte c'est que nous sachions qu'un événement parmi un million s'est réalisé, et que nous connaissions lequel. De même, la richesse de la phrase « la Française 2.31.10.67.482.305 est morte cette nuit » ne dépend pas du fait que cette victime m'est inconnue, ou qu'elle est, au contraire, ma femme.

De toute façon, il y a infiniment moins d'information dans cette phrase que dans le message contenu dans une seule image entrevue sur l'écran du télérécepteur. Tout simplement parce qu'il n'y a qu'environ 48 000 000 de messages du type « faire-part de décès », et que le nombre d'images télévisées possibles est si grand qu'il n'est pas trop de centaines de milliers de chiffres (au lieu de huit environ) pour le représenter.

Pratiquement, donc, qu'il nous suffise de savoir qu'un message est d'autant plus riche qu'il est improbable (1 chance sur 48 000 000 pour le premier message envisagé, 1 sur 1 000 000 pour le tirage de la loterie susmentionnée; en admettant ici — pour ne pas entrer dans de trop longs détails techniques — que tous les numéros sont probables et également tous les Français à égale merci d'une mort prochaine).

Reste à savoir comment codifier un message, et, surtout, comment évaluer l'information qu'il contient.

coder et chiffrer l'information contenue dans les messages les plus simples

Les plus simples des messages sont formés de mots ou de chiffres. Et il est naturel de dire que leur contenu d'information est proportionnel à leur longueur.

On rejoint ainsi l'idée liant probabilité et information : un message formé de quatre chiffres précise quelle éventualité s'est réalisée parmi 10 000 (10^4) nombres; un message de huit chiffres fixe un événement parmi 10^8 nombres.

Plus précisément, 10^8 est le carré de 10^4 :

$$10^8 = 10^4 \times 10^4,$$

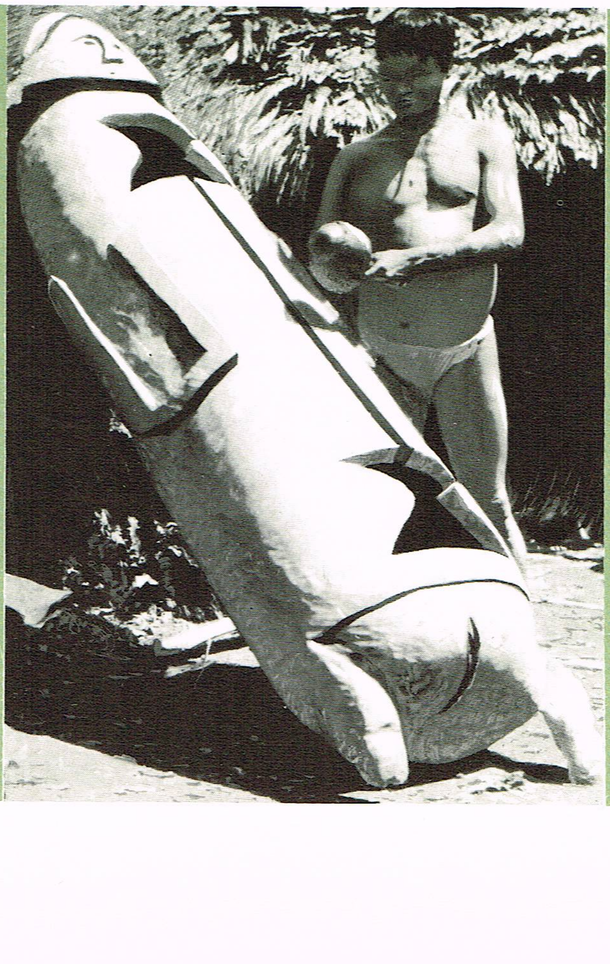
ce qui peut s'écrire

$$8 = \log\ (10^8) = \log\ (10^4) + \log\ (10^4) = 4 + 4.$$

Il suffit de poser **I = log N** (I étant la quantité d'information contenue dans un message fixant laquelle de N éventualités également probables s'est produite) pour donner une définition précise de l'information contenue dans certains messages simples, en accord avec la règle que l'on s'était imposée : « L'information contenue dans plusieurs messages est la somme des informations contenues dans chacun d'eux. » (En particulier, I est proportionnel à la longueur du message.)

Nous n'avons considéré que des messages chiffrés. Il existe d'autres possibilités : les messages alphabétiques, alphanumériques (contenant lettres et chiffres),...

Il est facile de chiffrer l'information contenue dans une suite de N caractères. Il y a, dans l'alphabet Morse, 41 « lettres » (26 lettres, 10 chiffres, 5 codes : « point », « erreur », « début », « fin », « intervalle »). Une « lettre » Morse constitue un message susceptible de prendre 41 états, et qui contient donc log 41 informations; un message de N lettres contient N log 41 informations en vertu du théorème précédent. Remarquons qu'un message formé de deux lettres prises parmi les quarante et une peut prendre 41^2 états : AA, AB, ...ZZ;



Transmettre l'information, problème majeur de toujours auquel on apporte des solutions chaque fois plus efficaces.

Phot. X.

un message de trois lettres, 41^3 états; de N caractères 41^N états. Ce dernier représente bien :
$I = \log (41^N) = N \log 41$ informations.

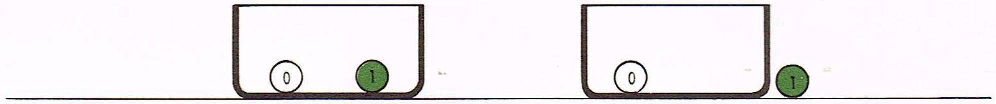
l'unité d'information

L'alphabet Morse a pour objet de coder un message écrit avec 41 « lettres » de manière à n'utiliser que deux symboles élémentaires : une brève et une longue (en fait, il y a un troisième signal utilisé : l'intervalle séparant les lettres ou les mots. Mais nous en ferons abstraction ici; il suffirait que le 41e symbole fût lui aussi codé en termes de brèves et de longues).

Beaucoup d'instruments physiques à traiter l'information — et surtout ceux qu'on appelle les « machines électroniques » — décomposent de même tout message en une série de caractères élémentaires qui ne sont susceptibles que de deux états.

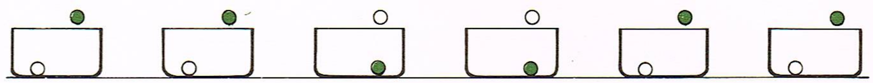
A cela une raison technologique simple : il faut utiliser des organes électroniques, seuls capables d'un fonctionnement ultra-rapide (de l'ordre de la microseconde, voire la nanoseconde [10^{-9} s]), alors que les organes mécaniques ou électromécaniques ont une inertie plusieurs centaines de milliers de fois supérieure. Or, il se trouve que la plupart de tels organes fonctionnent par « tout ou rien ».

Une information a été acquise.



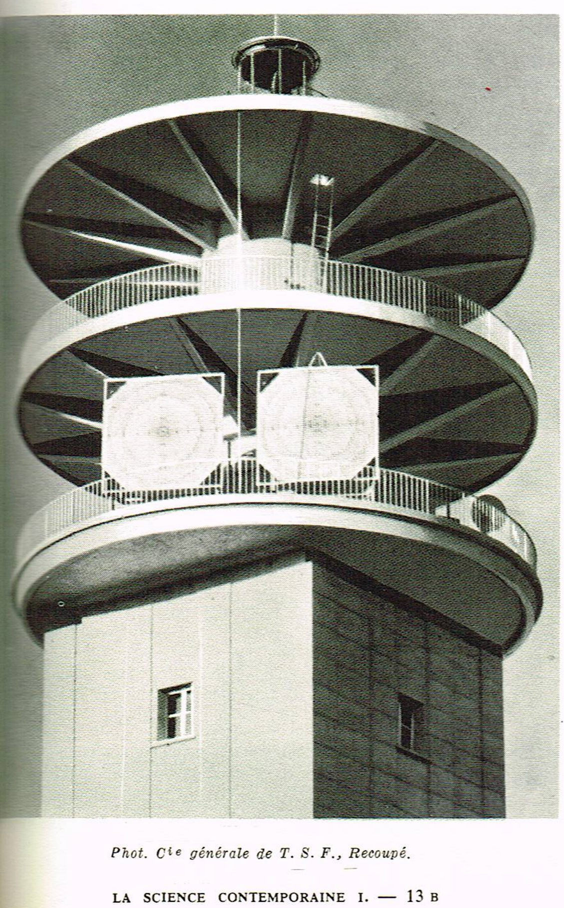
Un système élémentaire typique : tirage dans une urne contenant deux boules identiques, notées respectivement 0 et 1 (ou peintes en blanc et vert, ou comportant tout autre mode d'identification).
Etat initial : ignorance complète du résultat du tirage (probabilité *a priori* : 1/2 qu'il soit 0, 1/2 qu'il soit 1).
Etat final : nous savons que 1 est tiré (probabilité *a posteriori* : 0 que 0 soit tiré = certitude ou probabilité 1 [certitude] que ce soit 1).

Six informations acquises.



L'événement (110011) s'est réalisé :
10^6 états possibles, tous équiprobables (tirage équitable)
$\log. 10^6 \approx 19$ informations acquises.





Phot. Cie générale de T. S. F., Recoupé.

Plus précisément, les composants électroniques élémentaires ne sont susceptibles de prendre que deux états (nous verrons longuement dans le second tome, et notamment en mécanique quantique, que c'est le cas de tout système élémentaire physique ; la loi de quantification est générale, et toute quantification peut finalement se décomposer en loi de « tout ou rien ») :

— sous tension ou déchargés;

— électrisés positivement ou négativement;

— circuit ouvert ou fermé.

De plus, les mathématiciens ont inventé au siècle dernier une logique (algèbre de Boole) qui convient admirablement à l'analyse de tels organes. Il suffit d'appeler « vrai » un de ces états, « faux » l'autre, pour retrouver les règles de la logique la plus courante (règles tellement simples que leur étude mathématique a pu être conduite extraordinairement loin, et constituer une des branches les plus riches et les plus fructueuses des mathématiques modernes); de même il suffit de désigner les deux états respectivement par 0 et 1 pour retrouver l'arithmétique binaire des calculateurs électroniques.

Observons un système physique (élémentaire) susceptible de prendre deux états. Nous acquérons une unité d'information lorsque, à l'ignorance totale de son état, nous substituons la connaissance exacte de l'état qu'il prend.

Précisons la définition précédente : ce n'est pas le logarithme à base 10 qu'il faut prendre, mais celui à base 2, pour que l'information acquise par cette expérience élémentaire soit égale à $I = 1$ ($= \log 2$).

Un message nous fixant lequel de N événements s'est réalisé nous apporte $\log_2 N$ informations.

En toute rigueur, il faudrait préciser que les N états sont également probables. Apprendre qu'un événement rigoureusement certain s'est réalisé n'apporte guère d'information. Nous verrons dans le tome II comment généraliser aussi la notion d'information dans les cas les plus divers.

dAURJYT6T54RFDEJKICLMP9KUI87HYTG
BVCFR4EDXSZ3WXCVBHYTGF6REDFG5VHYf

Un message codé : 65 caractères (y compris codes « début » et « fin ») indépendants.

65 $\log_2$ 41 $\sim$ 348 informations

J'ARRIVERAI LE 14 JUILLET A LA
GARE MONTPARNASSE A 16 HEURES 45

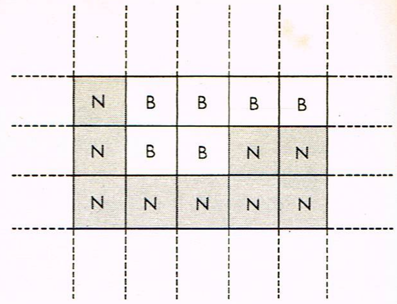
Un message en clair de 65 caractères, aussi coûteux à transmettre que le message précédent.

65 $\log_2$ 41 $\sim$ 348 informations

En fait, les caractères ne sont pas quelconques : après J'ARRIV il faut s'attendre à voir un E, et non un W. On dit qu'il y a **redondance** du message. Et il est connu qu'on peut transmettre un message beaucoup plus condensé aussi riche de sens pour son destinataire, mais beaucoup moins coûteux et de bien moindre information (au sens scientifique donné ici de ce terme). Par exemple :

ARRIVERAI 14-7 MTPARNASSE 16H 45



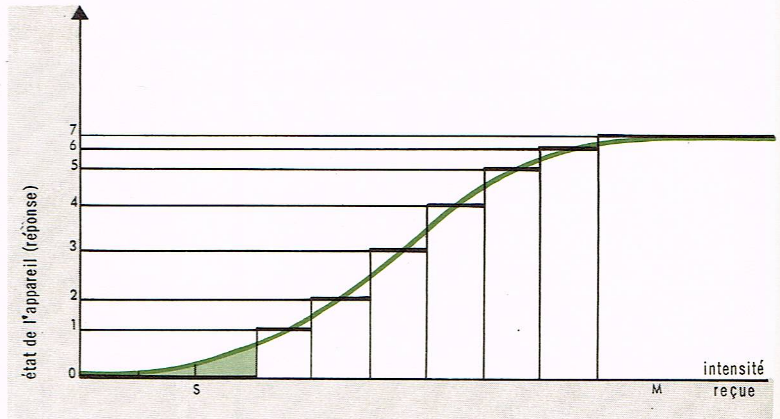


Un détail de l'image : 15 points (sur les 622 440 que comporte la scène).

Combien d'informations dans cette image de la télévision ?

Tout message reçu par un instrument physique ou un organe des sens, tel que l'œil ou l'oreille, peut se décomposer pareillement, et son contenu informationnel se chiffrer. L'image n'est pas continue : elle se décompose exactement en 819 lignes de 760 points, soit 622 440 cases. Si l'on admet que chaque case n'est susceptible que de 2 états (blanc ou noir), l'image est, en fait, un message NBBBB... de 622 440 caractères :

I = 622 440 informations

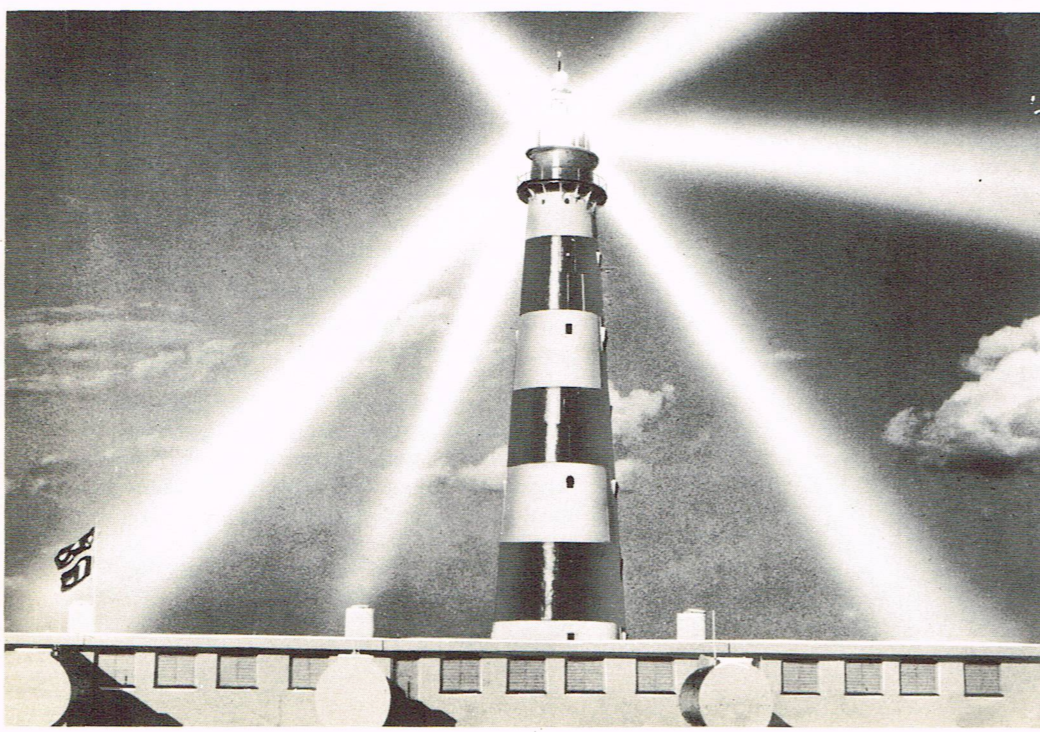


Il semble que le problème soit plus compliqué encore : chaque point peut être plus ou moins noir, l'intensité lumineuse pouvant varier de 0 (noir) à l'infini.

En fait, il n'en est rien : il y a un *éclat limite M* (au-delà duquel les lampes grilleraient) et un éclat minimal S *(seuil de sensibilité)*. De plus, deux intensités trop voisines donnent la même réponse *(seuil différentiel)*. *Finalement, toute réponse n'est susceptible que d'un nombre fini N d'états (loi de quantification* propre à tout système physique). Le message « portrait de speakerine » (de la Télévision française) comporte donc :

622 440 log N informations
(ici N = 8 **I = 1 867 320 informations)**

Bien entendu, il y a une énorme redondance, et une fiche d'identité judiciaire, qui tiendrait en quelques lignes, en dirait autant — à défaut de faire un égal plaisir — au spécialiste du portrait-robot. Là n'est pas le problème.



Feux et phares : première application des ondes électromagnétiques à la transmission de l'information.

Phot. J. de Beaupré et B. B. T.

l'information se transmet

Il importe de bien remarquer que l'information tire avant tout son intérêt de la grande facilité avec laquelle elle se transmet, c'est-à-dire voyage dans l'espace et dans le temps. Les sociétés préhistoriques ne connaissaient que la parole et le son, probablement pas les messages. Aussi les échanges étaient-ils très restreints, et les progrès stagnants. Certaines inventions remarquables (imprimerie et poudre, découvertes, dit-on, par les Chinois) sont restées ignorées pendant des millénaires, quand elles n'ont pas été perdues. La tradition orale fut longtemps le véhicule de la pensée-information, stockée dans les cerveaux avant de l'être dans les livres (et aujourd'hui dans les microfilms).

Il semble bien que l'information, unique moyen d'échange entre les pensées des hommes, se trouve à la base du développement des civilisations. L'espèce humaine, qui ne diffère pas fondamentalement des sociétés animales quant à la nourriture, la vie et la reproduction, s'en distingue profondément par la quantité d'information échangée entre ses membres. **On peut très bien mesurer un taux de civilisation par le volume annuel d'information transmise.** (Nous nous garderons toutefois de dire que la France est restée un pays barbare, d'après le nombre d'abonnés au téléphone, qui la place — je crois — au dix-septième rang dans le monde. Peut-être les Français, à défaut de leurs compagnes, plus bavardes au téléphone, sont-ils peu redondants, ou peu soucieux de communiquer.)

Une étude de l'histoire selon ces remarques reste à faire; elle serait d'un intérêt sociologique certain. Le physicien, lui, ne peut manquer de réfléchir sur les causes qui sont à l'origine de l'aisance — toujours relative évidemment — avec laquelle on peut transmettre l'information.

Cette possibilité participe d'un phénomène beaucoup plus général, et à vrai dire surprenant, qui est celui de la transmission à distance des états physiques (propagation). Ainsi le sol, la chaleur, la lumière, les champs électriques, magnétiques, la gravitation se propagent, et en général très loin. (Cette propriété banale est une des constantes de la vie courante; c'est pourquoi nous n'y prêtons pas attention. On pourrait toutefois imaginer un monde très rigide, à haute densité d'électrons, sans air, où ces phénomènes seraient très atténués.) Malgré les difficultés auxquelles l'explication de ces phénomènes de propagation a donné lieu, acceptons-en l'existence comme une donnée immédiate.

En quoi consistera alors le problème des télécommunications? Si nous savons faire dépendre d'une manière commode un état physique de l'information que nous voulons transmettre, nous l'aurons résolu en ce qui concerne le premier aspect de la question, qui est celui de la modulation. Moduler, c'est appliquer à un système physique approprié, dont l'état varie en conséquence, l'information de base (parole, écriture, image ou indication d'un appareil de mesure). C'est ainsi que la voix se traduit par des modifications successives, toutes liées entre elles, de la pression de l'air, de la résistance électrique des grains de charbon contenus dans le micro, de la variation de la tension et de l'intensité du courant électrique circulant dans les fils, du champ magnétique de l'aimant situé dans l'écouteur, de l'état de vibration de la plaque métallique, de la vibration de l'air situé entre l'écouteur et le tympan, finalement de celle des membranes et osselets de l'oreille interne. Et que, dans une caméra de télévision, l'image optique module l'intensité d'un faisceau d'électrons, qui commandera à son tour un « courant porteur » radio à haute fréquence, etc.

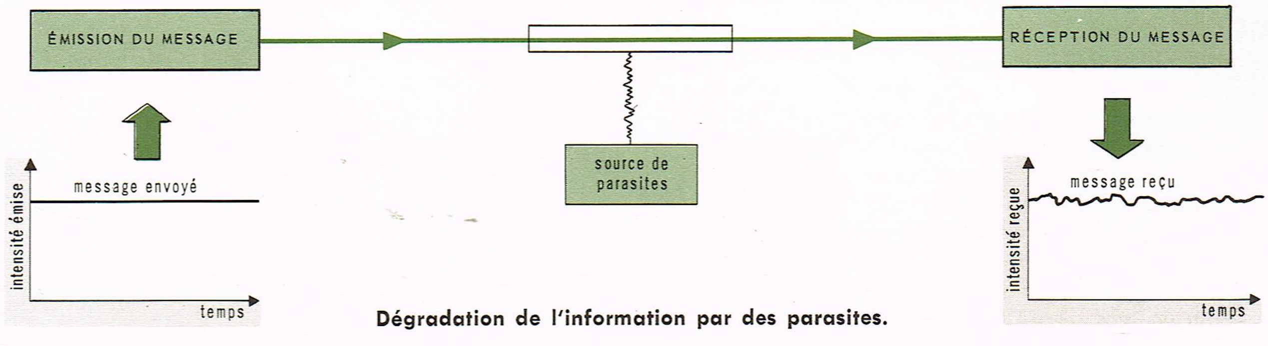
Reste le second aspect, qui est celui du choix du milieu de transmission. On s'efforce d'utiliser les milieux les moins dispersifs et les moins dissipatifs; ce qui signifie que l'on retient avant tout les milieux qui propagent les signaux avec un minimum de pertes et un maximum de concentration (directivité).

Ces milieux nous sont d'ailleurs imposés lorsqu'ils sont naturels : l'air (et le vide) essentiellement. Les recherches en vue d'utiliser la mer (liaisons avec les sous-marins) et la terre (transmissions par le sol, très peu vulnérable) sont toujours d'actualité. Les milieux artificiels se réduisent aux métaux bons conducteurs (aluminium), dont sont constitués tous les câbles.

Enfin, si l'air était autrefois le seul milieu physique que l'on sût moduler (parole, tam-tam), le champ électromagnétique s'est assuré aujourd'hui la quasi-exclusivité des transmissions par son incomparable facilité à subir les modifications et son énorme vitesse de propagation (son : 340 m/s; champ électromagnétique : 300 000 000 m/s).

elle se transmet mais elle se dégrade

Constatation banale en information : **on reçoit toujours autre chose que ce que l'on a envoyé** — « friture » dans le téléphone, parasites et bruits de fond en radio, points blancs (ou noirs) en télévision, craquements sur un électrophone, erreurs dans les télégrammes, rumeur se mêlant à la voix de l'orateur dans une salle, taches lumineuses sur l'écran d'un radar. Ces accidents gênants, qui mutilent l'information (jusqu'à la détruire lorsqu'ils atteignent une forte proportion), proviennent de deux causes principales.
Les signaux ont d'abord ramassé des « parasites » en ligne ou dans l'air. Les orages, les poussières sur les disques, les parasites d'allumage des moteurs, les coupures ou affaiblissements sur les câbles



Dégradation de l'information par des parasites.

sont autant de causes extérieures de perturbation. Ils donnent ce qu'on appelle des « bruits externes ». Ils ont comme particularité essentielle de pouvoir être évités, avec de plus ou moins grandes difficultés il est vrai. Le bruit qui affecte l'information de direction, de changement de vitesse ou d'assiette d'une « DS 19 » (à-coups de la pompe haute pression qui débite le liquide hydraulique) peut être réduit par l'augmentation de la vitesse de rotation de la pompe, par des ajustages calibrés qui régularisent le débit et agissent comme des filtres, par un soin particulier apporté au dessin des soupapes, par la multiplication des cylindres. L'utilisation d'un joint souple entre vilebrequin et arbre de transmission diminue le « bruit » du moteur, qui se traduirait par un fonctionnement saccadé (ralenti).

les inévitables bruits de fond, témoins de l'agitation thermique des électrons

Ce n'est, hélas! pas tout. Supposons un électrophone neuf, un disque bien propre. L'audition semble parfaite. Amenons à son maximum la puissance du son. Un grésillement (**bruit de fond**) apparaît, qu'on percevrait d'ailleurs avec la même intensité en arrêtant la rotation du disque. Le même inconvénient se retrouve en radio, sur l'écran de télévision placé à l'abri des parasites industriels et urbains. Ce bruit, interne, est dû à la structure granulaire de l'électricité. Au mouvement d'ensemble des électrons, régi par les tensions appliquées, et qui forme le courant électrique, se superpose une micro-agitation désordonnée, tout à fait analogue au fourmillement brownien. La température est responsable de ce mouvement anarchique; ou plutôt c'est ce va-et-vient incessant et brouillon qui constitue le phénomène « température », rigoureusement de la même manière que pour l'air : plus il est chaud, plus les molécules d'oxygène et d'azote s'agitent. **On comprend que ce bruit de fond, inhérent à la structure même de la matière, soit essentiellement inévitable.** Il existe dans tous les étages d'un poste de radio, mais le bruit thermique des lampes d'entrée est prépondérant; à l'instar du signal reçu par l'antenne, il traverse toute la chaîne d'amplification du poste, alors que le bruit engendré dans les lampes de sortie n'est soumis qu'à l'amplification de ces dernières.

Dans les récepteurs très sensibles, où le niveau du signal à l'entrée ne dépasse pas tellement celui du bruit thermique (radars, faisceaux hertziens, radiotélescopes, récepteurs de satellites), on s'efforce d'amplifier en travaillant à basse température (ce qui prohibe l'utilisation des lampes qui « chauffent »), dût-on pour cela travailler dans l'hélium ou l'azote liquide pour réduire aux limites extrêmes l'agitation néfaste des électrons (maser amplificateur).

Dans l'impossibilité de recevoir une information pure, force nous est d'en évaluer la qualité. Cela se fait tout naturellement par la mesure du rapport du signal au bruit (S/B), c'est-à-dire du quotient de la puissance du signal utile (musique, parole, etc.) par celle du bruit qui l'accompagne. Un tel rapport est évalué en décibels, par extension de l'utilisation d'une vieille unité d'acoustique, le décibel : rapport signal-bruit (en décibels) = 10 log décimal (S/B).

L'utilisation des logarithmes est purement conventionnelle, mais elle est commode : une conversation téléphonique est de très bonne qualité avec un rapport S/B de 50 décibels (puissance du signal = puissance du bruit × 100 000), de qualité moyenne entre 30 et 40, mauvaise au-dessous de 25 dB. La suite 50, 40, 30, 25 est plus aisée à manier que les rapports de puissance correspondants : 100 000, 10 000 1 000 et 300. On remarquera combien l'oreille est sensible à la gêne causée par le bruit; ainsi, une communication est jugée médiocre lorsque le bruit atteint le millième du signal (S/B = 30).

On comprend pourquoi la réduction du bruit à des niveaux très faibles par rapport à l'information utile est un problème si important et si difficile.

Après avoir établi le caractère scientifique de l'information, nous avons appris à la mesurer, puis remarqué qu'elle se transmet assez aisément, mais qu'elle se dégrade; il faut maintenant aborder les problèmes pratiques posés et essayer de montrer comment ils ont été résolus.

le prix à payer, chiffré par la largeur de la bande passante

Nous connaissons tous le prix, le volume et le poids d'une chaîne de radio haute fidélité. Le fin mot de l'histoire est extrêmement simple. L'oreille humaine perçoit normalement les sons s'étageant des très graves (120 Hz) aux très aigus (15 000 Hz). Il est d'ailleurs rare qu'un son émis soit pur. (Ce n'est jamais le cas d'une note de musique, quel que soit l'instrument : on dit que la note est accompagnée d' « harmoniques », dont la richesse constitue le timbre. La fréquence de vibration se nomme aussi hauteur du son.) C'est avant tout le timbre qui différencie les instruments : piano ou violon, un

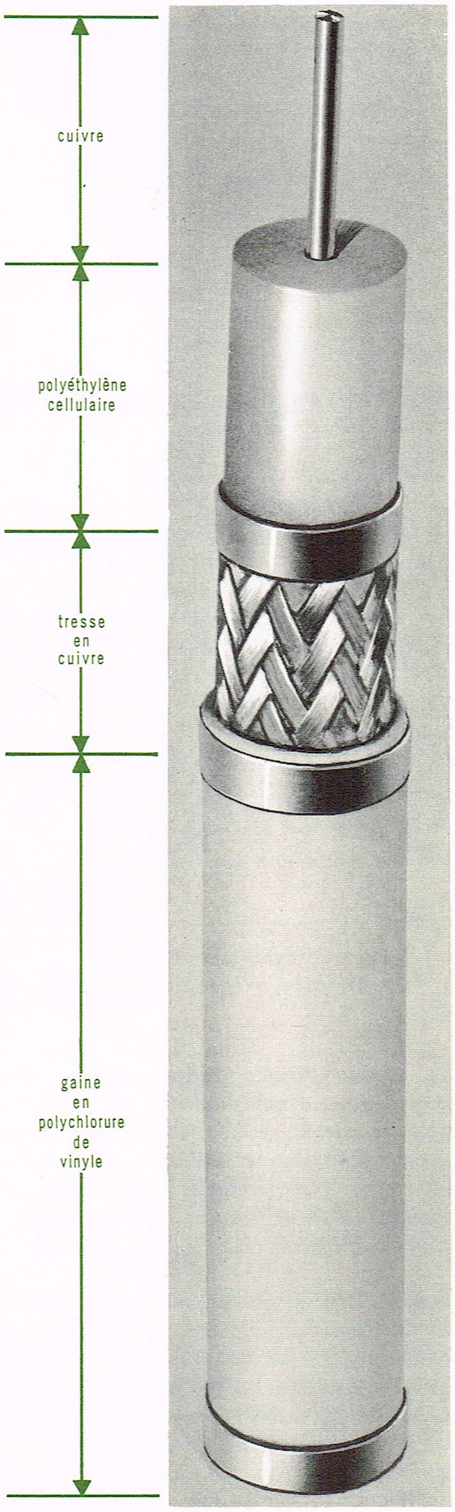


Une conversation téléphonique est de très bonne qualité avec un rapport signal sur bruit de 50 décibels.

Phot. Refot-Rapho.

« sol_3 » reste un « sol_3 ». Dire que le la_3 de votre piano est riche en harmoniques 2, 3 et 4, cela signifie simplement que votre Pleyel ou votre Steinway, après frappe convenable, émet la vibration 435 (la_3 fondamental), accompagnée des vibrations plus faibles de fréquence 870 (435 × 2), 1305 (435 × 3), 1740,... (harmoniques d'ordre 2, 3, 4). Cela n'est pas l'effet du hasard : il est physiquement impossible, sauf précautions extrêmes, qu'une corde à piano, un boyau de violon, un corps de saxophone, ou une plaque métallique émettent un son non accompagné d'harmoniques. Il en est de même de la voix humaine. Quand vous entendez une symphonie — ou une conversation —, un grand nombre de sons frappe l'oreille à chaque instant, chacun muni de ses harmoniques. Tout se passe comme si, à chaque instant, vous « entendiez » une grande quantité de fréquences différentes. L'analyse de toutes les fréquences qui composent un son, aussi compliqué soit-il, s'appelle analyse « spectrale » ou « harmonique », opération très analogue à la décomposition de la lumière par un prisme. Une telle analyse joue un rôle si fondamental en physique moderne (c'est un élément de quantification) qu'un chapitre entier du second tome lui sera consacré. **Finalement, la bande passante, ou largeur de bande, mesure l'étendue de la gamme de fréquence — l'étendue du spectre — qu'un récepteur peut reproduire sans déformation.** Chacun a constaté que les voix deviennent graves au téléphone, que les récepteurs habituels de radio affaiblissent ou suppriment même les sons aigus; ils prêtent aux discours une sonorité grave — ce qui est d'ailleurs de bon ton si l'on songe à la nature habituellement peu plaisante des informations dont ils nous rassasient. C'est qu'un appareil téléphonique « coupe » à 3 400 Hz, un poste de radio à 5 000-7 000 Hz, autrement dit que les sons de fréquence supérieure sont atténués ou étouffés. Un bon électrophone est fidèle jusqu'à 11 000 périodes, la haute fidélité reproduit sans déformation tous les sons audibles et monte à 16 000 périodes... et à des prix très élevés. Car cela suppose que chaque étage d'amplification est capable de traiter, sans les altérer, une grande étendue de fréquences. On conçoit qu'il est plus simple de se contenter d'une gamme étroite. **La consommation en bande passante est onéreuse.** Et, de même que les sensations auditives disparaissent au-dessus de 17 000 périodes (les ultrasons ne sont pas entendus), tous les supports matériels de transmission sont affectés des mêmes limitations : un ensemble de deux fils conducteurs (dont un pour boucler le circuit) transporte fort bien le courant continu (fréquence zéro), mais n'est plus bon à rien au-dessus de quelques dizaines de kilocycles, car il affaiblit trop les fréquences élevées, surtout sur de longues distances.

Voilà pourquoi la télévision (fort gourmande), qui exige 10 mégahertz (10 millions de périodes) de bande passante, ne peut utiliser les fils comme le téléphone, même pour une courte descente d'antenne. On leur préfère le câble coaxial, formé d'un conducteur au centre d'un cylindre métallique rempli d'un diélectrique (polyéthylène) qui sert au « centrage » et favorise la propagation. La théorie du coaxial est un peu difficile à expliquer ici; contentons-nous de remarquer que ces câbles guident très bien le courant — plus exactement le champ électromagnétique — jusqu'à des fréquences très élevées (plusieurs mégahertz, voire même des milliers sur une courte distance). Quant à l'atmosphère, sa bande passante est énorme, mais limitée à partir des fréquences de 20 000 à 30 000 mégahertz par l'absorption due d'abord à la vapeur d'eau, puis aux molécules d'oxygène ou d'azote qui entrent en résonance à ces très hautes fréquences.





Ce qui coûte cher, c'est la bande passante.

Phot. Jan Klein et Paul Genest.

la bande passante, mesure de la capacité d'information

On conçoit que plus la bande est large, plus important est le volume d'information qu'elle peut transmettre. Il est manifeste que l'auditeur est mieux informé d'un concert par un très bon électrophone que par l'intermédiaire d'un poste datant de 1920, à bande très étroite. A l'extrême, un canal qui ne saurait transmettre qu'une fréquence ne porterait aucune information, puisque la présence continuelle du même signal ne signifie rien.

En fait, **la capacité d'information d'un moyen de transmission est proportionnelle à sa bande passante.** Cette capacité se mesure en unités d'information par seconde, donc en logons (ubi ou bits) par seconde. On se doute bien que cette capacité est détériorée lorsqu'un fort bruit accompagne l'information.

C'est ce qu'exprime la formule très simple due à Tuller et Shannon :

$$C = L \log_2 \left(1 + \frac{S}{B}\right)$$

(C, capacité, en unités/seconde; L, largeur de bande, exprimée en périodes; S et B, intensités du signal et des bruits).

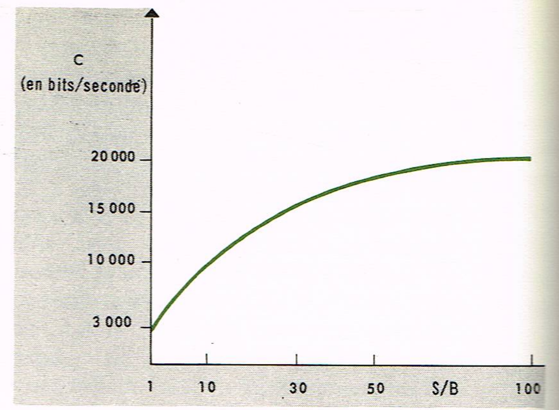
Le rapport signal/bruit de la voie de transmission est exprimé comme un rapport de puissance. Le logarithme à base 2 est une vieille connaissance. **Le débit d'information croît avec la bande passante et la force du signal; il est fonction décroissante du bruit.** Il s'agit bien de débit, et l'on peut très bien comparer un débit d'information au débit d'eau d'une canalisation, d'autant plus important que la section est grande. Il est dangereux de pousser trop loin les analogies, mais, pratiquement, la puissance du signal est comparable à la pression de l'eau, qui accentue son écoulement, les frottements et les obstacles représentés par les aspérités du tuyau jouant le même rôle que le bruit.

Prenons un circuit téléphonique. S'il obéit aux règles standards internationales dites « normes C. C. I. T. T. » (Comité consultatif international des téléphones et télégraphes), sa bande passante s'étend de 300 à 3 400 Hz. Cette largeur a été choisie comme déformant suffisamment peu la voix (1).

La capacité d'information d'une bonne ligne peu bruyante (S/B = 10^5, correspondant à 50 dB) vaut : $C = (3\,400 - 300) \log_2 (1 + 10^5) = 50\,000$ unités/seconde.

En télévision (largeur de bande : 10 mégahertz = 10^7 périodes), on aurait, avec un rapport S/B de 30 dB,
$C = 10^7 \log_2 (1 + 10^3) \sim 100$ millions d'unités/seconde.

(1) Une telle norme est généralement respectée sur les grandes artères du niveau P. T. T., appelées « lignes à grande distance » (L. G. D.). Elle ne l'est pas dans de nombreux circuits locaux, de caractère moyenâgeux (à Paris, par exemple).



Capacité d'information (C), en fonction du rapport signal/bruit (S/B), pour la bande téléphonique de 3 100 cycles.

Or, nous avons vu que, *grosso modo*, l'information d'une image de télévision représente 1 million d'unités. La R. T. F. nous distribue 25 images à la seconde. Elle écoule donc 25 millions d'unités par seconde et utilise le canal TV non loin de sa capacité maximale. On notera l'énorme disproportion des capacités entre les canaux téléphoniques et ceux de la télévision (rapport 2 000). Il est heureux que la taxe de télévision ne respecte pas cette proportion comme l'Administration pourrait l'imposer d'après un critère scientifiquement valable.

quelques procédés modernes de transmission de l'information

Le premier moyen utilisé pour se faire comprendre a certainement été après le geste la parole, réduite à l'articulation de sons simples servant à désigner des objets courants. (On imagine difficilement le néandertalien se lançant dans de longues périodes. Encore moins le sinanthrope.)

Nos ancêtres se devaient d'être peu redondants (ce souci s'est perdu depuis), tant leurs moyens de communication étaient réduits : portée limitée à quelques centaines de pas selon la force du gosier et le vent, vitesse limitée à celle du son (340 m/s). La signalisation par feux fut un grand progrès, atténué par le fait que l'on savait assez mal s'en servir pour converser. C'est qu'on ne savait pas encore « coder » l'information comme on le sut beaucoup plus tard avec le télégraphe Chappe. Puis vinrent l'écriture et, plus tardivement encore, l'imprimerie. Avec la création d'un service des postes, la vitesse de propagation est passée de celle du coureur de Marathon à celle des chevaux de poste, et enfin à celle des chemins de fer. Au XIX[e] siècle, le télégraphe et le code Morse furent inventés, et avec eux apparurent les énormes vitesses de transmission, allant jusqu'à 300 000 kilomètres par seconde, que permet l'onde électromagnétique. (L'auditeur de mondovision entend un concert donné à 5 000 km avant même le spectateur installé dans la salle où joue l'orchestre.)

Aujourd'hui, le trafic d'information qui intéresse le monde moderne appartient à quatre types :
— transmission de la parole, ou téléphonie;
— transmission de l'écriture, ou télégraphie;
— transmission des images, ou télévision;
— transmission des « données ».

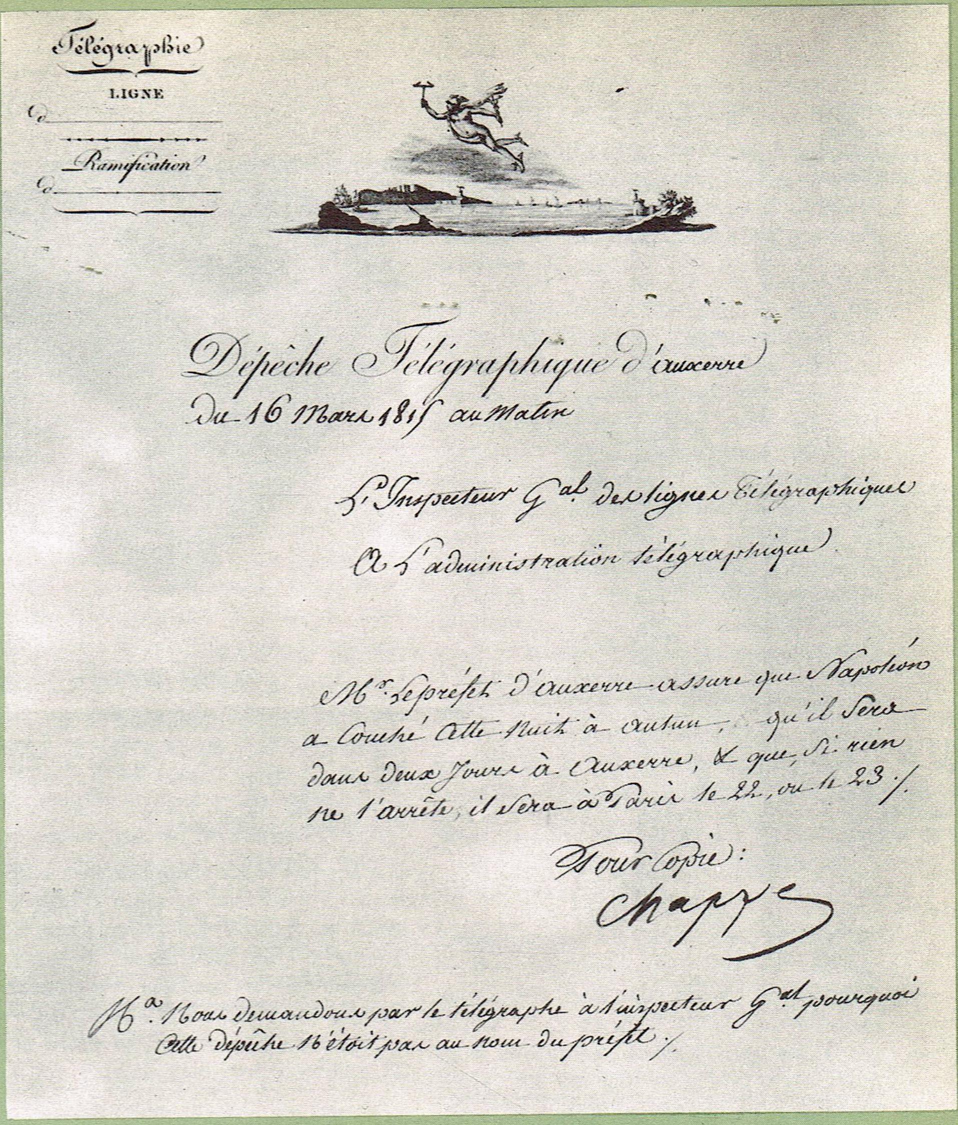
Cette dernière rubrique, d'apparition toute récente, vient du fait que, après l'homme, les machines se sont mises à leur tour à

« parler ». Ainsi conversent les machines comptables d'une grande banque avec l'agence centrale, les calculateurs de défense aérienne qui se « passent » des pistes d'avions, l'émetteur de télémesure d'un satellite avec les « terriens ». Comme ces machines à calculer et ces équipements automatiques sont très bavards, bien que peu redondants, ils ont rapidement débordé les possibilités de la télégraphie classique, et c'est ainsi qu'on a rangé leur information dans une catégorie à part, spécifique : les « données ».

les techniques téléphoniques : courants porteurs et commutations

Tout le monde connaît le principe du téléphone, qui utilise la modulation d'un courant obtenue par la pression de la voix.

Chaque poste d'abonné est loin d'être utilisé vingt-quatre heures sur vingt-quatre. Réserver une ligne fixe entre chaque abonné et son correspondant est évidemment impossible. Les voies sont en fait « banalisées » par l'emploi de la « commutation ». Celle-ci peut être manuelle — une opératrice établit elle-même le circuit — ou automatique. Le réseau P. T. T. — *grosso modo* — est constitué de grandes voies omnibus (les lignes à grande distance : L. G. D.) qui desservent les réseaux locaux d'abonnés, les nœuds de ce système d'artères étant occupés par les antennes de commutation.



Télégraphie
LIGNE
d
Ramification
d

Dépêche Télégraphique d'auxerre
du 16 Mars 1815 au Matin

L'Inspecteur Gal. des lignes Télégraphiques
A l'administration télégraphique

Mr. le préfet d'auxerre assure que Napoléon a couché cette nuit à autun, qu'il sera dans deux jours à auxerre, & que, si rien ne l'arrête, il sera à Paris le 22, ou le 23./

Pour copie :
Chappe

Na. Nous demandons par le télégraphe à l'inspecteur Gal. pourquoi cette dépêche n'étoit pas au nom du préfet./

Dépêche télégraphique Chappe. On apprenait à coder l'information.

Or, les câbles coaxiaux utilisés par les L. G. D. ont une bande passante bien supérieure à celle de la voie téléphonique normale, limitée par accord international à la bande 300-3 400 périodes. On a donc imaginé, pour profiter au mieux de leur capacité, de leur faire écouler simultanément 12, 60, 120... voire 960 voies téléphoniques. C'est le procédé de « multiplexage » à courant porteur.

Il existe une opération électronique, appelée « transposition », qui consiste à décaler d'une même quantité toutes les fréquences d'un signal. Dans ces conditions, une voie téléphonique transposée, au lieu d'occuper le spectre 300-3 400, occupera celui de 4 300-7 400 (décalage 4 000), 8 300-11 400 (décalage 8 000), etc. Cette transformation est d'ailleurs simple, et n'exige que des filtres électriques soignés pour couper les fréquences indésirables qui apparaissent. Il reste à juxtaposer par transposition les spectres des 12, 60 ou 960 voies téléphoniques, de sorte qu'elles restent distinctes et qu'on puisse, par l'opération inverse, les séparer à la réception. L'opération obéit à des règles internationales, afin que le passage d'un pays à l'autre n'oblige pas à changer la transposition.

Illustration page 214.

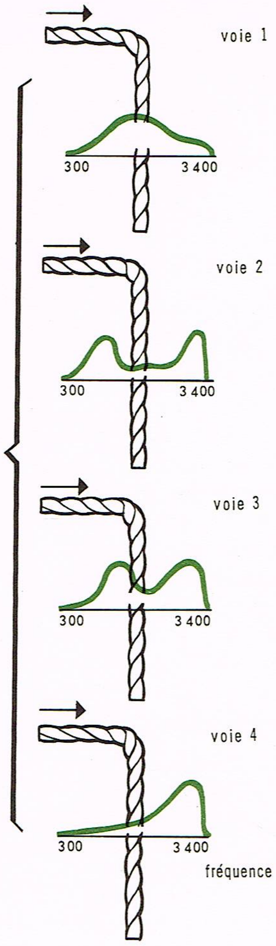
les techniques télégraphiques : le téléimprimeur

Point n'est besoin de parler du code inventé par Morse en 1844, et toujours utilisé. Un autre code s'est depuis imposé à l'universalité, c'est le code dit « à 5 canaux » (alphabet C. C. I. T. T. nº 2). C'est le langage des téléimprimeurs et de tout l'appareillage télégraphique associé. Il consiste à représenter une lettre, un chiffre ou un signe orthographique par cinq impulsions de courant. Chaque impulsion peut prendre deux valeurs, selon qu'elle est positive (+ 48 volts) ou négative (— 48 volts). L'alphabet ainsi constitué serait à 32 signes ($2^5 = 32$) et ne permettrait pas de représenter les 26 lettres, les 10 chiffres et les autres codes. La difficulté a été tournée par le même procédé que celui qui est employé sur la machine à écrire, où la frappe d'un chiffre s'opère en tapant un signe de ponctuation, après avoir appuyé sur la touche « majuscule ». Dans le code à 5 canaux, il existe deux symboles réservés à l'inversion « chiffre-lettre », qui avertissent le récepteur que le signe qui suit chacune de ces combinaisons est un chiffre ou une lettre, selon que la

Phot. Larousse (Arch. nat.).

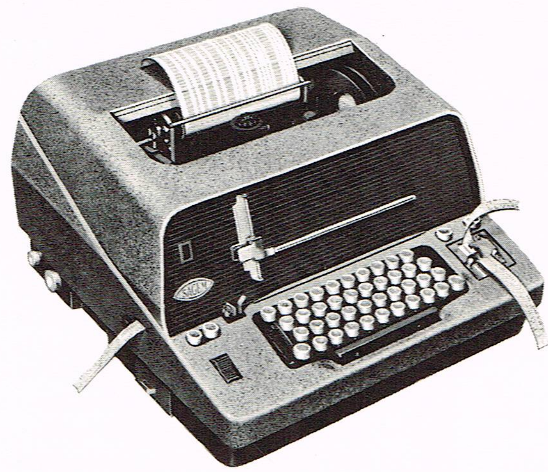
Représentation schématique de la transposition et du multiplexage.

VOIES INDIVIDUELLES



Au lieu d'utiliser 4 voies téléphoniques individuelles de 3 100 hertz...

... on se sert d'une seule ligne et on décale les fréquences de trois voies afin de juxtaposer les quatre messages dans une bande de 15 400 hertz.



Téléimprimeur transistorisé S. P. E.

On distingue la bande portant des inscriptions perforées en code à 5 canaux.

marque qui précédait était relative à une lettre ou à un chiffre. D'ailleurs, toutes les spécifications des téléimprimeurs sont calquées sur celles de la machine à écrire, par la nécessité bien simple de disposer d'un clavier universel sans d'inextricables difficultés d'adaptation.

L'utilisateur d'un téléimprimeur frappe sa lettre comme sur la plus classique des machines à écrire. La frappe d'une touche déclenche un mécanisme d'impression (permettant de contrôler visuellement la frappe) et l'émission d'une série de cinq impulsions (+ ou —) d'une « largeur » de 20 millisecondes. Pour séparer ces signes et donner à l'organe mécanique de réception le temps d'interpréter les signaux reçus, chacun d'eux est précédé d'un signal de 30 ms, appelé « départ », et suivi d'un autre signal de 20 ms, appelé « arrêt ».

On voit que l'émission d'une lettre dure en tout 150 millisecondes. Si l'on faisait abstraction des codes « arrêt » et « départ », le téléimprimeur pourrait envoyer 50 impulsions par seconde. (Cependant, la cadence réelle d'émission d'un téléimprimeur est liée à la vitesse de la dactylographe. Elle est donc irrégulière.)

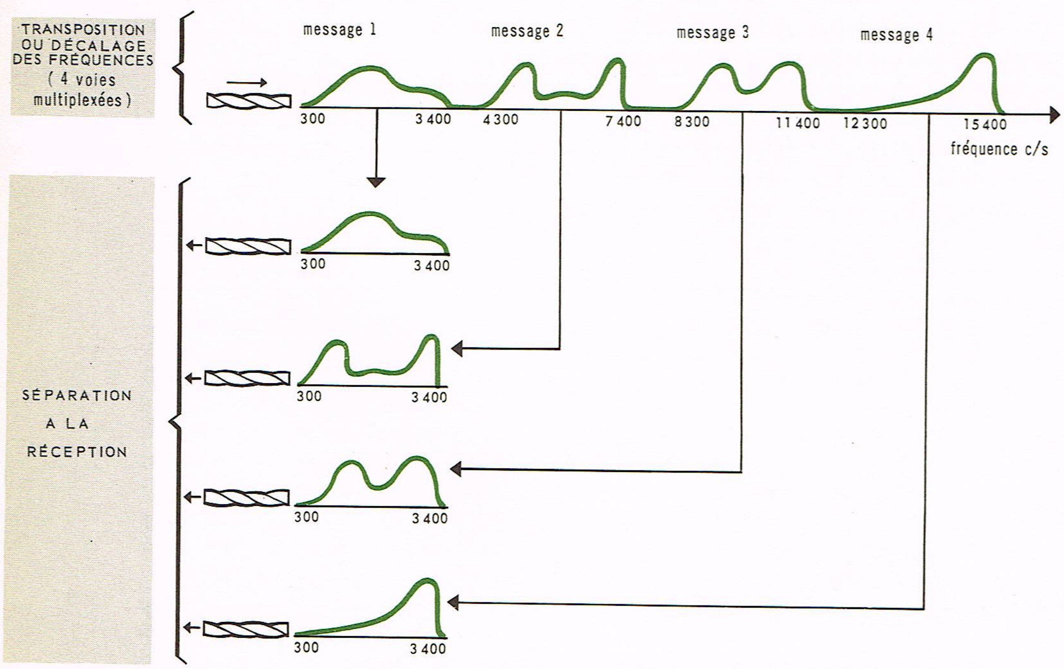
La largeur de bande réservée à l'écoulement de ce trafic est normalisée à 120 périodes. Elle est très confortable, puisqu'elle permettrait théoriquement, avec un rapport S/B de 10 dB, d'écouler plus de 360 impulsions par seconde :

$$C = 120 \log_2 (1 + 10) \sim 400.$$

Rien n'empêche d'utiliser la technique des courants porteurs en télégraphie, et de juxtaposer par exemple douze voies qui peuvent être transmises simultanément sur le même canal : c'est la « télégraphie harmonique ».

L'information à 5 canaux est avantageusement transcrite sur une bande perforée, étroit ruban de papier où chaque lettre est représentée par la perforation ou la non-perforation de cinq trous.

L'intérêt de la bande perforée est qu'elle peut être ensuite lue automatiquement à cadence régulière — comme un film. Des palpeurs ferment les contacts lorsqu'ils rencontrent un trou, et la transformation « code perforé → courant en ligne » s'effectue aisément. L'appareil de lecture et d'émission de bande est le transmetteur automatique. Enfin la bande est un excellent moyen de stockage.



Phot. SAGEM.

savoir communiquer avec les machines

Comme il vient d'être dit, le trafic de conversations entre les machines — calculatrices diverses, engins téléguidés, etc. — constitue un flot très particulier d'informations et ne peut plus, de nos jours, être traité comme la télégraphie classique. Les raisons précises de ce privilège sont au nombre de deux. Si l'on peut admettre généralement sans grand inconvénient des fautes et erreurs dans un télégramme, il est absolument exclu d'en accepter dans une comptabilité (machines bancaires), dans une mesure de position d'engin ou de température de tuyère. **Les transmissions de données doivent être très sûres.** Enfin, la rapidité de fonctionnement des machines est très supérieure à la célérité d'une dactylo. Les émetteurs-récepteurs de données seront adaptés à de grandes vitesses, très supérieures à 50 unités/seconde.

Très répandus aux Etats-Unis, les équipements de transmission des données sont encore relativement peu commercialisés en France, où leur utilisation concerne surtout les forces armées. Ainsi, les calculateurs de défense aérienne se communiquent, sous forme d'un flux ininterrompu de données, les trajectoires des avions reconstituées automatiquement à partir des échos du radar, élaborant ainsi une situation aérienne globale. Le flot des informations déborderait une voie télégraphique ordinaire. Aussi utilise-t-on une voie téléphonique, dont la capacité théorique atteint 30 000 unités/seconde (pour un rapport S/B de 30 décibels). En fait, seules les vitesses de 750 ou de 1 500 unités/seconde sont employées, la qualité réelle des équipements et le développement de la technique interdisant actuellement d'aller beaucoup plus vite.

Il va de soi que le code utilisé n'est plus forcément le code international à 5 canaux. C'est, en l'occurrence, le code binaire qui est presque universellement employé, avec des variantes, dans les calculatrices. Dans ce code, on n'utilise que deux chiffres (0 et 1), au lieu des dix du code décimal auquel nous sommes habitués.

On voit que le code binaire exige davantage de chiffres pour écrire un nombre — à peu près trois fois plus —, mais cet inconvénient est amplement compensé par le fait qu'il n'utilise que deux signes — ce qui « plaît » beaucoup aux circuits des machines à calculer —, dont chacun est susceptible de prendre seulement deux états — l'état ouvert ou l'état fermé —, tout comme les interrupteurs. **C'est donc parce que l'électronique « aime » bien travailler par « tout ou rien » que le système binaire est d'emploi commode.**

Sur la ligne de transmission, ces 1 ou ces 0 se traduiront, par exemple, par la présence ou l'absence d'un niveau courant, par la présence ou l'absence d'une fréquence. Pour diverses raisons, la transmission des données dont nous avons parlé utilise deux fréquences. Le 1 est représenté par la fréquence de 1 125 périodes, le 0 par celle de 1 875. (Quand on ne reçoit rien, c'est que la ligne est coupée...) Chaque « impulsion » dure 1,33 milliseconde, puisqu'il y en a 750 par seconde.

On pourrait objecter que la largeur de bande nécessaire est quasi nulle, puisqu'il n'y a que deux fréquences pures à transmettre. Cela est vrai, sauf aux points de raccordement, où l'on passe brusquement d'une fréquence à une autre. Cette discontinuité exige de la bande passante, de même qu'un coup de bélier se propage bien dans un tuyau de large section et se trouve étouffé dans une petite canalisation. On pourrait montrer que les signaux formés de 0 et de 1 équivalent à des sommes de fréquences multiples de la fondamentale 375 ($1\,125 = 375 \times 3$; $1\,875 = 375 \times 5$; $750 = 375 \times 2$) jusqu'aux ordres les plus élevés. Le spectre du signal est un « spectre de raies ». On conçoit qu'il faille que le canal physique transmette fidèlement, sans déformation excessive, le signal.

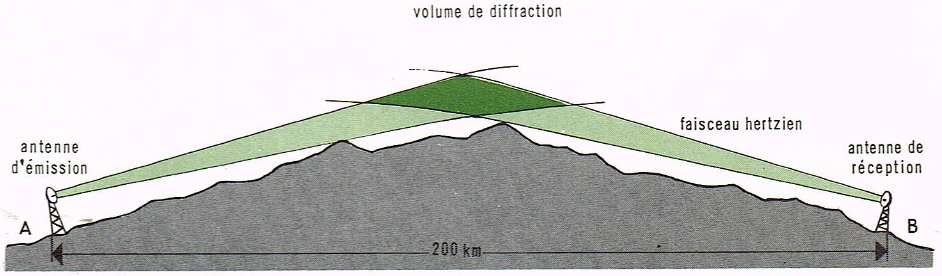
Les erreurs, avons-nous dit, ne sont pas tolérables. Mais elles sont cependant inévitables, même si elles n'affectent pas couramment un chiffre sur dix mille. Ce taux est pourtant trop élevé pour un chef comptable et, quelquefois, pour une position d'avion ennemi. Or, l'information concernée n'a pratiquement pas de redondance : on ne peut pas « corriger les fautes d'orthographe » quand il s'agit de nombres. On se tire de ce mauvais pas en ajoutant de l'information supplémentaire volontairement et selon des règles connues du récepteur. On pourra, par exemple, découper la suite de 1 et de 0 en tranches de cinq chiffres et ajouter à chaque tranche un 1 ou un 0 de façon que chacune d'elles possède un nombre pair de 1.

Ainsi, la suite de 15 chiffres
10010 11011 10011 devient
100100 110110 100111.

L'équipement récepteur doit trouver un nombre pair de 1 tous les six chiffres. Sinon il y aura une erreur, un nombre impair de 1 ou de 0 s'étant transformés en 0 ou en 1 : il devra demander une répétition. On dit qu'il y a « détection d'erreurs ». On s'aperçoit d'ailleurs qu'on ne les détecte pas toutes (par exemple, une double erreur, changeant deux chiffres, ou permutant l'ordre de deux informations du même groupe). On peut augmenter également l'efficacité du contrôle : il suffit d'ajouter encore de l'information. Certains systèmes permettent même de détecter et de trouver la place des erreurs : ce sont les codes correcteurs. Mais, bien sûr, ils rallongent la longueur, donc le coût, du message.

les faisceaux hertziens troposphériques

Nous connaissons tous les faisceaux hertziens. Ce sont des émetteurs-récepteurs radio qui se distinguent de l'équipement de radiodiffusion par quelques traits. Leurs antennes (souvent du genre parabolique) et leur fréquence élevée (au-dessus de 50 MHz, donc au-delà des ondes courtes; la longueur d'onde va de 8 m à 3 cm) en font des émetteurs de radio très directifs. Au lieu d'éparpiller dans l'espace leur information, ils la concentrent dans



Propagation troposphérique des ondes hertziennes.

une seule direction, d'où le nom de « câbles hertziens » qu'on leur donne parfois.

Cette directivité entraîne une telle économie de puissance, que l'on peut diffuser sur plusieurs canaux d'information (12, 60, 120, 960 voies téléphoniques). Les faisceaux hertziens transportent de l'information multiplexée. Les relais hertziens de la télévision sont bien connus.

Tous les faisceaux hertziens dits « classiques » travaillent à vue directe, les stations étant à visibilité optique l'une de l'autre. La portée est de l'ordre de 50 km en moyenne, et varie selon les conditions géographiques. Or, depuis 1934, et grâce notamment à Marconi, on a constaté qu'on pouvait recevoir des signaux, évidemment très atténués, à des distances allant jusqu'à plusieurs centaines de kilomètres. Comme les techniciens d'aujourd'hui savent faire des antennes de grande dimension (antennes à grand gain), des récepteurs très soignés et peu bruyants, des émetteurs à forte puissance (l'unité de puissance est ici le kilowatt, et non plus le watt comme dans les équipements à vue directe), les énormes pertes constatées sont supportables.

Le phénomène de réception des signaux hertziens lointains ne s'explique pas par la réflexion sur des couches atmosphériques ionisées, comme c'est le cas pour les ondes courtes, et il s'explique mal par la diffraction. Force fut donc d'admettre qu'il s'agissait d'un processus nouveau, localisé dans les basses couches de l'atmosphère, d'où le nom de « diffusion troposphérique » donné à ce nouveau type de propagation. En fait, le grand responsable est l'indice de réfraction de l'air, certes toujours très voisin de 1, mais variant très vite d'un point à un autre, selon l'état local des vents, des tourbillons, et la teneur en vapeur d'eau.

L'énergie envoyée par l'antenne A se propage toujours en ligne droite. Mais une infime fraction (par exemple 1/100 000) est diffusée, diffractée, réfléchie dans toutes les directions à cause de l'hétérogénéité des masses d'air. L'énergie qui est envoyée dans la bonne direction est reçue en B. Il est facile de donner une image de la propagation troposphérique. Dans une pièce bien propre et fermée, un rayon de soleil trace un pinceau lumineux invisible à l'œil qui n'est pas sur son trajet. Faisons de la poussière ou de la fumée de tabac. Grâce à la diffusion de la lumière par les fines particules, le rayon devient visible de toute la pièce.

Le scintillement de la mer provient de ce qu'entre l'observateur et le soleil il existe une infinité de petites surfaces d'eau planes et changeantes qui, à un instant donné, forment miroir. Il en est de même en propagation troposphérique lorsque l'air est calme et peut être considéré comme une juxtaposition plus ou moins fluctuante de feuillets à indice de réfraction constant.

Il convient enfin de noter que l'intensité reçue au cours d'une liaison troposphérique est toujours affectée de variations énormes, pouvant atteindre un facteur 1 000 d'un moment à l'autre, un peu à l'instar du phénomène expliquant le « fading » à la réception. C'est une caractéristique un peu analogue au fading de la radio. On combat ce grave inconvénient par l'emploi de la « diversité » : au lieu de recevoir sur un seul récepteur, on en utilise deux ou quatre. A chaque instant, par commutation, on conserve le signal venant du récepteur où il est le plus fort. On amoindrit ainsi grandement les variations du champ.

Tous ces faits déterminent la nature du matériel troposphérique et ses caractéristiques :

— puissance de 1 à 10 kW de 500 à 5 000 MHz (70 cm à 7 cm);

— antennes d'un diamètre allant jusqu'à 20 m;

— 4 récepteurs, 2 antennes, 2 émetteurs pour une station terminale en quadruple diversité.

Comme on ne peut parler que d'un signal statistique moyen, la qualité sera exprimée en pourcentage du temps. Par exemple, on dira que telle liaison de 150 km ou 250 km fonctionne correctement pendant 99,9 p. 100 du temps du mois le plus mauvais (en général pendant l'hiver).

L'intérêt de la diffusion troposphérique est manifeste. Elle permet, au prix certes d'une complication d'équipement, de multiplier par trois ou quatre la longueur des trajets hertziens. Elle est donc bien adaptée à une desserte extensive, aux terrains difficiles qui rendent très coûteuse l'infrastructure.

Aussi voit-on surtout cette technique utilisée dans les pays en voie de développement industriel. Et, bien sûr, par les forces armées de toutes les nations.



Relais pour la transmission de faisceaux aériens troposphériques.

Phot. C. S. F.

LES ORDINATEURS
au SERVICE DE L'INTELLIGENCE

par
Claude PICARD



Cette machine
de la faculté des sciences d'Orsay
pèse 16 tonnes et coûte 10 millions;
elle est capable d'effectuer
15 millions d'additions
par minute.
La résolution d'un système
de cent équations du premier degré
à cent inconnues,
qu'un opérateur exercé
mettrait quinze ans à effectuer,
lui prend 16 secondes!

Ces chiffres donnent une idée du foudroyant essor de cette branche de la connaissance qu'est l'informatique, née vers 1922. Les possibilités des ordinateurs (« machines électroniques » ou « calculateurs ») ont décuplé tous les trois ans de 1953 à 1965, ce qui correspond à un facteur de dix mille en douze ans. Et les experts proclament : les prochaines années verront naître des machines d'une technologie révolutionnaire, aux performances difficilement soupçonnables, à côté desquelles le matériel actuel apparaîtra comme préhistorique.

Comment sont donc constitués ces appareils aux possibilités si vastes? Comment travaillent-ils? Quels sont les services qu'ils peuvent rendre? Sont-ils vraiment indispensables? Quel avenir possèdent-ils et quels seront leurs effets sur la science et sur la civilisation?

C'est à de telles questions que nous aimerions répondre, sans employer une terminologie difficile et complexe, mais en insistant sur quelques problèmes généraux et en faisant revivre cette passionnante aventure de l'automatisation mise au service de l'intelligence.

Phot. A. B. C.

La variété des connaissances nécessaires à la conception et à la réalisation des calculateurs électroniques, mais aussi le travail que doivent fournir de leur côté les utilisateurs ainsi que la diversité des domaines d'application font de l'informatique une science jeune et dynamique, qui absorbe dès maintenant une part considérable des activités humaines. Il est banal de dire que la civilisation des années 70 verra s'étendre le champ des loisirs, et que les bienfaits de l'automation n'y seront pas étrangers; il est au contraire rare d'entendre affirmer que le degré éminent de croissance des connaissances acquises par l'homme conduirait à un étranglement par manque d'informations, si la conduite des recherches et des applications n'était guidée par une méthode et une organisation scientifiques que seuls les calculateurs ont rendues possibles.

comment calculait-on auparavant ?

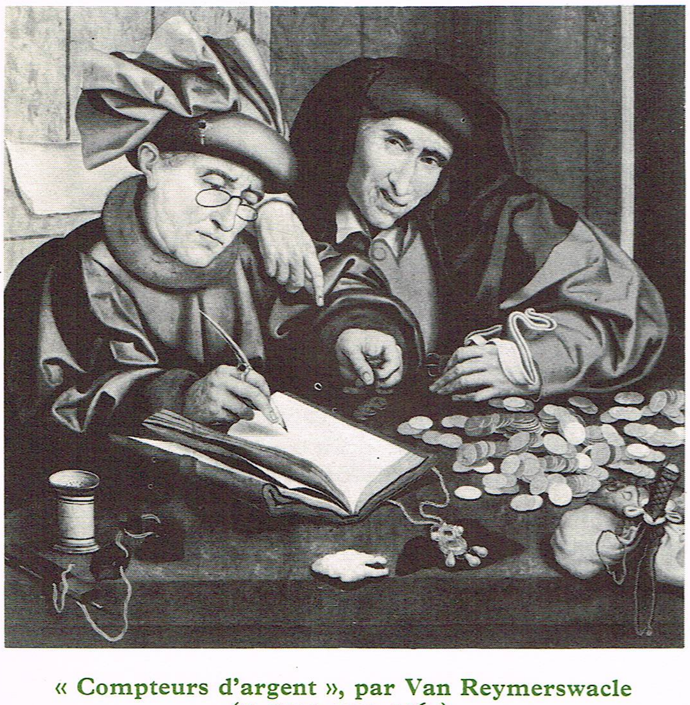
Il a fallu beaucoup de temps pour que l'humanité apprît à faire des calculs sur les nombres entiers et dégageât à ce sujet les notions fondamentales de l'algèbre moderne, telles que nous les verrons au chapitre *les Mathématiques de nos enfants* (tome II). Elle sut d'abord compter sur ses doigts, puis inventa le boulier.

Cet instrument n'est pas automatique, car il n'effectue pas le report auquel donne lieu l'addition 17 + 85. **C'est pour cela que l'invention par Pascal de sa machine peut être considérée comme une étape décisive :** une fois enregistré le premier nombre 17, il suffit de faire tourner la première roue de 5 positions et la deuxième de 8 pour qu'apparaisse le total 102. La première roue s'est déplacée de 7 + 5 = 12 positions, mais, lors du passage par 0, faisant suite au 9 et correspondant à 10, une impulsion mécanique a fait avancer d'une unité la roue des dizaines, déjà placée sur 1; et, de même, la rotation de cette deuxième roue de 8 autres positions a enclenché la roue des centaines.

D'immenses progrès techniques ont abouti aux machines électromécaniques, désormais familières à tous. Mais le principe en reste le même, et les limitations analogues : le temps de calcul est pratiquement égal à celui de l'affichage des données et des facteurs successifs. **Aucune mémoire n'est disponible, et il faut sans cesse recopier.** Pour effectuer

1 751.532.412 + 203.3 271

il faut afficher 1 751 et 532, noter le produit 931 532 sur une feuille, enfoncer les touches correspondant à 931 532 et à 412, noter le résultat : 383 791 184. Et de même pour 203 et 3 271. Il reste encore à ajouter le résultat précédent à ce nouveau produit pour obtenir le nombre cherché, soit 384 455 197.



« Compteurs d'argent », par Van Reymerswacle (v. 1500-apr. 1567).

Finalement, nous voyons que les opérations proprement dites sont faites extrêmement vite (en deux à trois secondes), alors que le temps d'affichage et d'écriture est relativement plus grand (30 secondes environ pour enfoncer 36 touches, y compris celles qui correspondent aux opérateurs . de multiplication et + d'addition.)

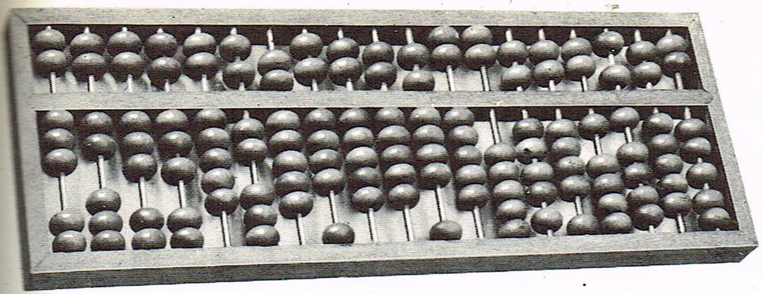
C'est en cela que le fonctionnement des machines électromécaniques diffère de celui du calculateur opérant à la main : pour ce dernier, il est beaucoup plus long d'effectuer les calculs que d'en mettre en place ou d'en aligner les facteurs. Et déjà l'on voit apparaître ce qui peut être énoncé à propos de toute étude sur les calculateurs électroniques, en cela semblables à leurs prédécesseurs électromécaniques : **calculer est aisé et rapide, stocker et mettre en place les données et les résultats extrêmement long et coûteux.**

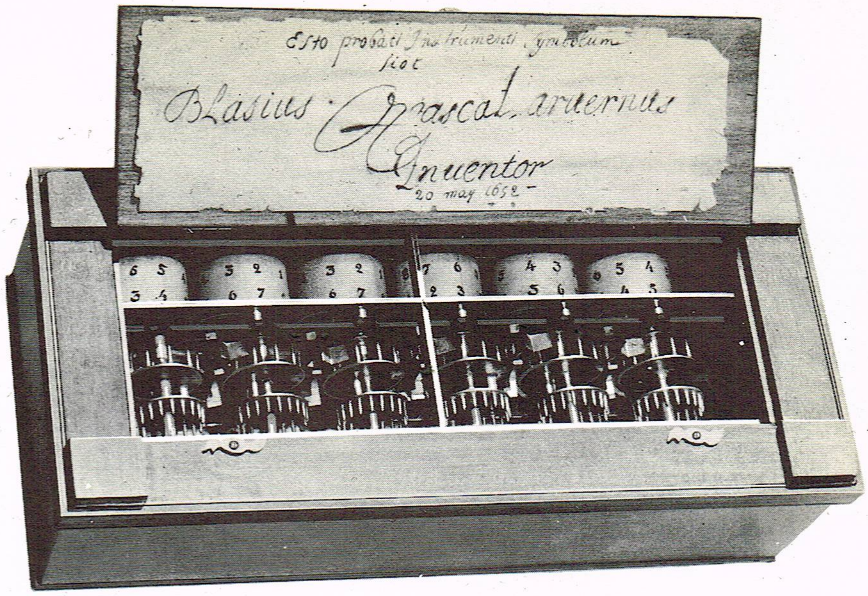
Cependant, les vrais ancêtres des calculateurs modernes sont les calculatrices électrocomptables à cartes perforées, encore exclusivement utilisées dans de très nombreuses entreprises modestes.

L'emploi de cartes perforées remonte à l'invention, il y a cent cinquante ans, par Jacquard, du métier à tisser avec lequel le dessin des tissus est préfiguré par les trous des cartons guidant les lisses agissant sur les fils. Puis, le mathématicien anglais Babbage (1792-1871), qui a consacré une partie de son activité d'astronome au calcul de tables et d'éphémérides, imagina et commença à réaliser une machine destinée à soulager l'homme de toutes les fonctions purement automatiques. Après la construction d'un calculateur réalisant certaines fonctions bien particulières avec huit chiffres exacts, l'inventeur conçut, sans arriver au bout de son projet, un calculateur universel comportant les cinq parties que l'on retrouve dans les ordinateurs : entrée par cartes perforées (empruntées à Jacquard), unité de contrôle mécanique travaillant avec des cartes, opérateur arithmétique, mémoire à roues dentées, enfin sortie avec impression mécanique. Malheureusement, ces vues géniales étaient en avance d'une centaine d'années sur les possibilités techniques de l'époque. Plus tard (vers 1880), l'Américain Hollerith imagina et breveta la façon moderne de coder l'information sur cartes, que ce soit pour l'impression ou pour la lecture.

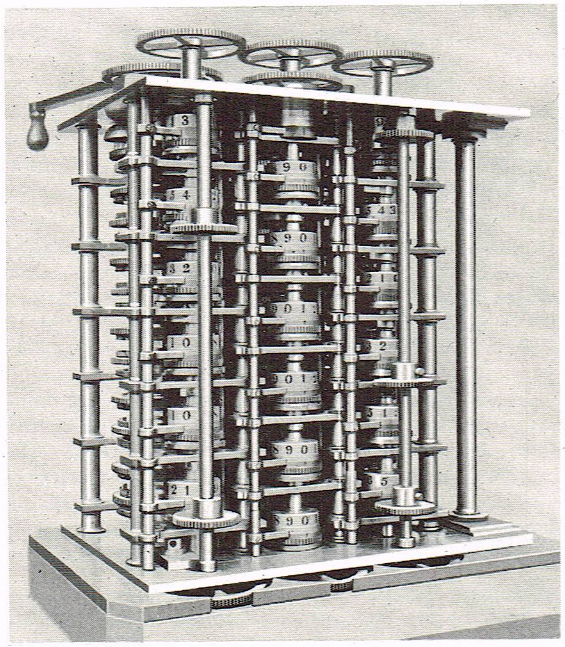
Il fallut attendre la fin de la Seconde Guerre mondiale pour que naquît la première calculatrice. Grâce à la conjugaison des progrès réalisés sur machines de bureau, machines à facturer, machines à registres multiples et tabulatrices à cartes perforées, la machine d'Aiken, connue sous le nom de « Harvard Mark I » (1944), avait des performances et une sécurité de fonctionnement qui paraissent aujourd'hui dérisoires, mais qui permettaient déjà des calculs atomiques fondamentaux. Vint tout de suite après l'ENIAC d'Eckert, la première machine électronique au monde, équipée de dizaines de milliers de tubes occupant tous les panneaux d'un très vaste local, à l'usage des spécialistes de la balistique. Plus tard, les constructeurs de

Musée des Beaux-Arts, Nancy; phot. Giraudon.





Quelques ancêtres : le boulier, la machine de Pascal *(ci-contre)*, celle de Babbage, et, *à droite*, la première tabulatrice Bull.





matériel à cartes perforées se sont intéressés particulièrement à l'électronique; c'est ainsi qu'aux Etats-Unis et en France sont apparues, vers 1950, les premières calculatrices commercialisées. Et chacun sait quels progrès étonnants ont été permis, en une vingtaine d'années, par les recherches dans le domaine de la logique et dans la technologie des composants électroniques et des supports de l'information.

QU'Y A-T-IL DANS UN ORDINATEUR ?

Lorsqu'on pénètre dans un laboratoire où fonctionne un grand ordinateur moderne, on voit qu'il est formé de beaucoup d'organes différents, interconnectés par un réseau très complexe de gros câbles discrètement cachés sous le plancher de la vaste salle qui le contient.

L'organe le plus vivant semble être le pupitre devant lequel travaille l'opérateur et où clignotent un très grand nombre de lampes; mais il y a aussi d'imposantes armoires contenant chacune une *bande magnétique* analogue à celle d'un magnétophone et se déroulant à une très grande vitesse.

Quelle est donc l'utilité de ces bandes?

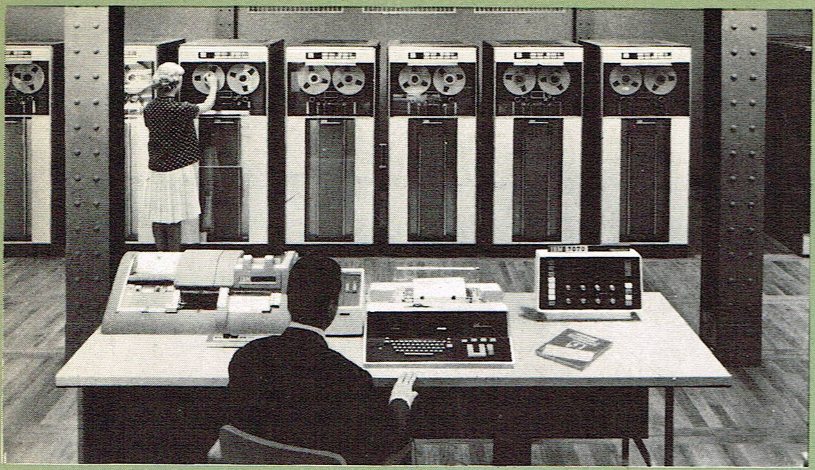
Ce sont des **mémoires,** c'est-à-dire des éléments susceptibles d'enregistrer un message, de le mettre en réserve (en vue d'une utilisation, d'une consultation ou d'une lecture ultérieures), indéfiniment ou jusqu'à ce qu'un nouveau **mot** soit introduit ou efface le contenu en question, et de le communiquer à l'organe de calcul, à un autre élément de mémoire (pour le transfert) ou encore à l'utilisateur (pour l'écriture).

Nous avons précédemment vu que l'absence d'une mémoire était précisément un des désagréments essentiels des machines électrocomptables. Cela se traduisait par la nécessité de recopier sans arrêt; ce qui est déjà fâcheux dans le calcul manuel, mais infiniment plus grave lorsqu'il s'agit d'une machine électronique qui peut travailler à une vitesse dix mille fois supérieure à celle de l'entrée ou de la sortie d'une information.

stocker l'information

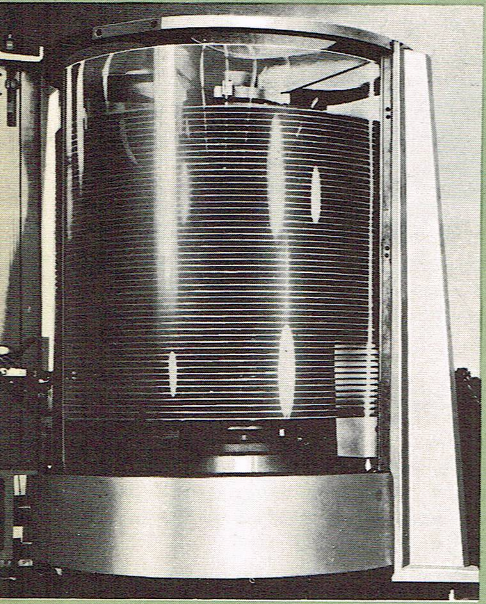
Sur la bande dont il a été question a été déposée une couche dont chaque élément est susceptible de prendre deux états magnétiques, conventionnellement appelés 0 et 1.

Chacun de ces éléments est capable d'enregistrer exactement une unité d'information (ou ubi), et un organe annexe — appelé « transcodeur » — permet de traduire auparavant toute information en écriture binaire, c'est-à-dire en n'utilisant que des 0 ou des 1. De telles bandes permettent de stocker une centaine de millions d'unités d'informations. Le déroulement est calqué sur celui du magnétophone, mais certaines précautions telles que le maintien de la tension par un procédé pneumatique sont à prendre, compte tenu de la très grande vitesse de lecture. En fait, la bande doit avancer des quelques centimètres correspondant à la lecture d'un ensemble donné d'informations, puis s'arrêter à l'endroit précis où se termine

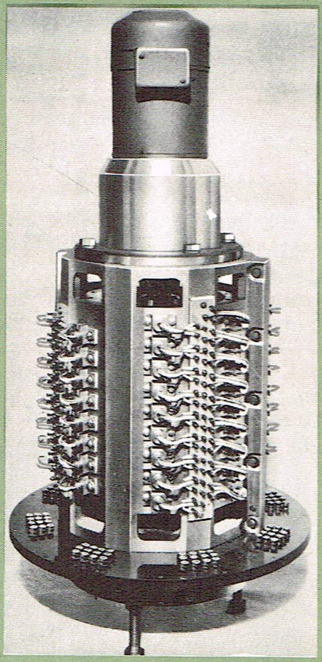


Un centre de calcul électronique. On distingue les bobines des bandes magnétiques.

TROIS ORGANES DE MÉMOIRE

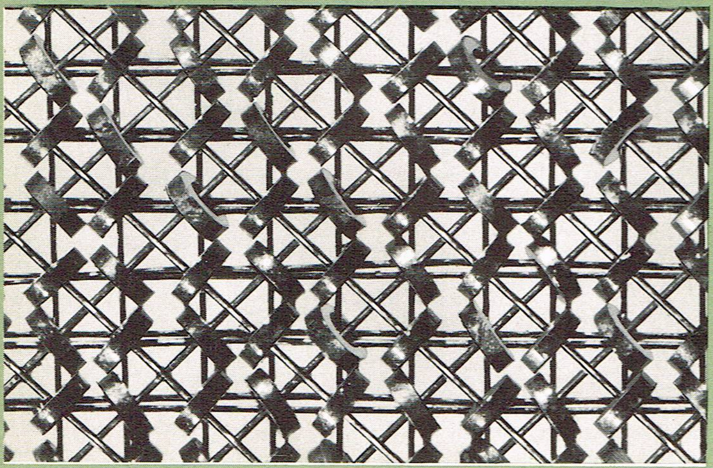


La pile de disques.



Le tambour magnétique.

Partie de la matrice à tores.



l'information à lire, et repartir dès que la machine fait de nouveau appel à sa mémoire, c'est-à-dire souvent au bout de quelques millièmes de seconde. Il n'est donc pas étonnant que de tels organes soient extrêmement délicats et coûteux. De plus, les informations doivent être stockées dans l'ordre où elles seront utilisées; la marche arrière ou la recherche d'un nombre placé n'importe où sur la bande peut immobiliser un ordinateur d'une valeur de plusieurs millions de francs en « attente lecture » pendant les quelques secondes nécessaires pour chercher un nombre situé très loin en avant ou en arrière.

Pour remédier à ce dernier inconvénient, on a pensé à utiliser des *disques magnétiques*. Une tête de lecture correspond à chacune des faces de la pile de disques, et un bras amovible permet de l'approcher plus ou moins de l'axe de manière à recueillir, ou à enregistrer, les informations sur l'une des pistes concentriques qu'ils portent. Le délai de recherche aléatoire de l'information est réduit au temps nécessaire pour faire tourner d'un angle convenable le disque et placer correctement la tête de lecture, temps qui est considérablement inférieur (il s'agit d'une fraction de seconde) au temps de déroulement d'une centaine de mètres de bande. Ainsi, les disques sont utilisés comme mémoire de grande capacité, susceptible de contenir des centaines de milliers de mots avec un temps d'accès moyen relativement court.

Assez comparables aux disques sont les *tambours magnétiques*, cylindres magnétisés comportant par exemple 128 pistes, chacune munie d'une tête de lecture. Si la capacité de chaque piste est de 8 192 (128 mots de 64 ubis), on peut enregistrer jusqu'à 1 048 576 (16 384 mots) informations; ce nombre est très inférieur à celui qui correspond aux bandes, mais le temps maximal de recherche d'une information mal placée est celui qu'il faut pour que le mot correspondant passe devant la tête de lecture, c'est-à-dire un tour de tambour, sensiblement 1/200 de seconde pour un tambour tournant à 12 000 tours à la minute. Les tambours sont utilisés maintenant comme mémoire de grande capacité et de temps d'accès moyen parce que leur prix de revient est relativement faible devant celui des systèmes très rapides.

Un autre type de mémoire très utilisé est formé d'une matrice de *tores de ferrite*, c'est-à-dire de petits anneaux magnétisables, connectés de telle manière que le temps d'accès moyen, de l'ordre du dix-millionième de seconde, soit compatible avec le temps demandé pour effectuer une opération arithmétique. Il s'agit donc d'une mémoire rapide, mais particulièrement onéreuse : chaque élément coûte quelques francs, et il ne peut donc exister sur une machine électronique qu'au plus quelques millions de tels éléments. Un ordinateur pourra comprendre aussi quelques dizaines de millions de tores de ferrite moins fins à temps d'accès vingt fois plus élevé, mais moins coûteux.

à l'avant-garde de la technique

Plusieurs types de mémoires ont été développés dans de nombreux laboratoires de recherches et possèdent des temps d'accès particulièrement intéressants. Citons les mémoires à films minces (utilisées dans « Univac 1107 » et « Atlas I »), les mémoires associatives, devant permettre la sélection des informations de manière radicalement nouvelle, et surtout les mémoires cryogéniques.

Ces dernières utilisent les phénomènes de supraconductivité des métaux à une température voisine du zéro absolu : la résistance électrique s'effondre alors brusquement, et le seuil de cette discontinuité est lié au champ magnétique auquel est soumis le métal; un fil de métal placé à une température de 4 °K (— 269 °C) présente une résistance quasi nulle en l'absence de tout champ. Le temps de commutation, c'est-à-dire de passage de l'état 0 (absence de courant dans un solénoïde; pas

Phot. I. B. M., Cie des Machines Bull, C. S. F.-Bouillot.

de champ, donc résistance nulle) à l'état I (solénoïde sous tension; résistance non nulle), ou réciproquement, est de l'ordre de 0,1 μs. Une cellule de mémoire ainsi réalisée par un couplage de deux éléments (supraconducteur + solénoïde) est appelée « cryotron ». La réalisation d'ensembles de grande capacité formés de telles mémoires pose naturellement des problèmes techniques ardus, dont la réalisation est liée aux techniques des très basses températures.

Pour nous résumer, disons que les éléments des ordinateurs (unité centrale et mémoire, organes d'entrée, de sortie et de liaison) dépendent : du mode de représentation des informations qui correspondent aux données, de l'ensemble des instructions disponibles et des techniques de réalisation. La mémoire d'un calculateur dont le rôle est de conserver le maximum d'informations et de transmettre le plus rapidement possible tout mot nécessaire à l'unité de calcul est caractérisée par deux critères : la capacité et le temps d'accès. Mais, en fait, en concevant la mémoire d'un calculateur il faut tenir compte également de raisons économiques, de sorte que l'ingénieur dispose d'un choix de mémoires qui permet de concilier au mieux capacité, vitesse d'accès et prix de revient. La rapidité des perfectionnements apportés à leurs applications permet de commercialiser et de rendre d'un usage courant les éléments — unité centrale aussi bien que mémoire — moins d'un an après que les recherches les concernant ont été menées à bien.

dix chiffres, ou seulement deux?

Toute mémoire est formée d'éléments susceptibles de prendre chacun deux états, c'est-à-dire que toute information doit être codée en alphabet binaire (voir le chapitre « Ce qu'est l'information », et notamment le paragraphe « Coder et chiffrer une information »). Nous verrons en effet que la technologie électronique et peut-être plus encore la structure logique d'un ordinateur imposent une telle codification.

Explicitons donc le système de numération correspondant, permettant de représenter tout nombre entier par une succession de 0 et de 1.

Le symbole décimal 1965 représente le nombre

$$1\,000 + 900 + 60 + 5,$$

ou encore $1 \times 1\,000 + 9 \times 100 + 6 \times 10 + 5 \times 1$,

ou $1\,(10^3) + 9\,(10^2) + 6\,(10^1) + 5\,(10^0)$,

et, de même, 10,27 représente le nombre

$$1\,(10^1) + 0\,(10^0) + 2\,(10^{-1}) + 7\,(10^{-2})$$

Eh bien, de la même manière, le nombre écrit en binaire

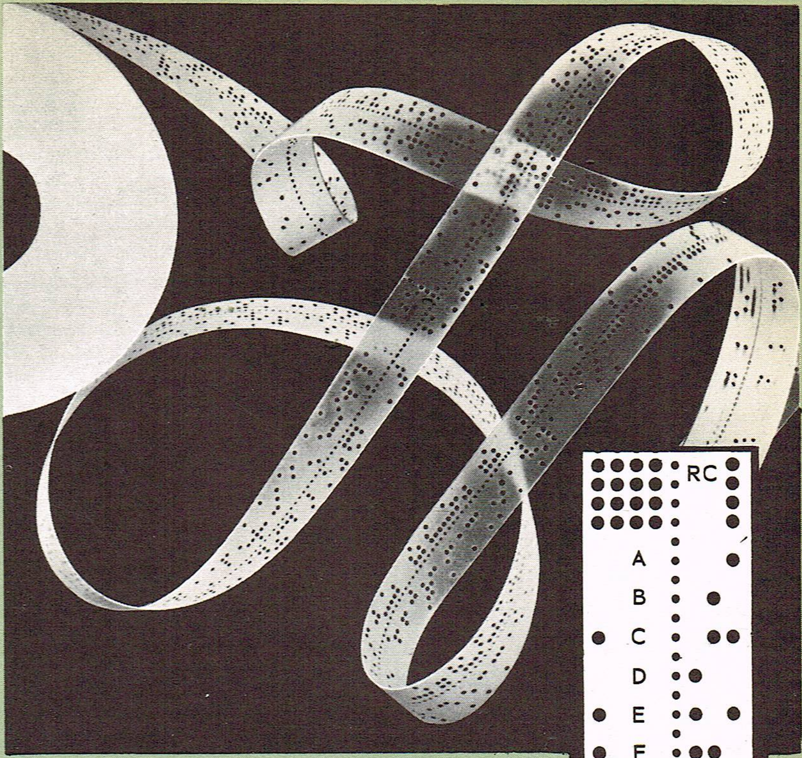
101100101

a pour valeur

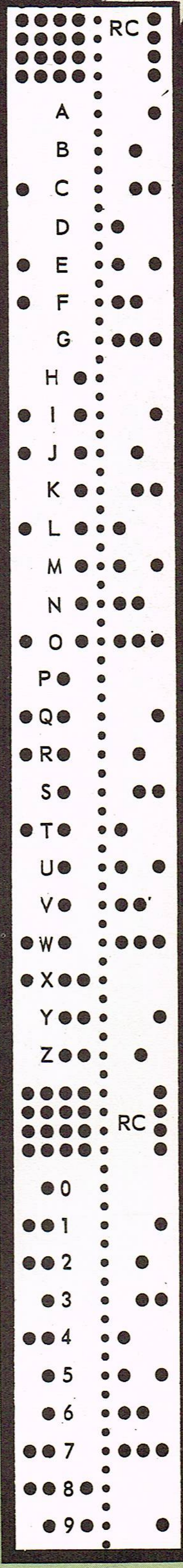
$1\,(2^8) + 0\,(2^7) + 1\,(2^6) + 1\,(2^5) + 0\,(2^4) + 0\,(2^3) + 1\,(2^2) + 0\,(2^1) + 1\,(2^0) = 256 + 64 + 32 + 4 + 1 = 357.$

Il est souvent utile aussi de traiter l'information alphabétique, et même des caractères spéciaux. Il est par exemple commode de désigner les opérations par un code parlant facilement à l'esprit, et il sera plus facile d'écrire M pour « multiplier » que 43 par exemple. De même, la machine doit savoir lire le nom des clients et reconnaître des signes tels que $<, >, \leqslant, \neq$...

Aucune difficulté à cela; il suffit de disposer d'un ensemble de 6 ubis susceptible de prendre $2^6 = 64$ états, et de convenir d'un code, arbitraire, portant sur 64 caractères au plus (par exemple 10 chiffres, 26 lettres, et 20 caractères spéciaux; il y aura alors 8 combinaisons de 0 et de 1 qui n'auront pas de signification, et que la machine n'aura jamais à traiter).



Phot. Cie des Machines Bull.



A chaque ligne de la bande perforée correspond un symbole international. Les 6 colonnes de droite permettent de représenter au plus 64 symboles (10 chiffres, 26 lettres, les quatre signes + ; / ..., ainsi que des codes divers, tels que « blanc », « point », « à la ligne », « contrôle d'imparité », etc. On remarque que le nombre de perforations est toujours impair (il y a une perforation en 7e colonne si — et seulement si — il y en a eu 2, 4 ou 6 dans les six autres). Lors d'un transfert d'une information, on vérifie cette parité.

Quelques systèmes de numération.

décimal	binaire	octal	décimal codé binaire
1	1	1	0001
2	10	2	0010
3	11	3	0011
4	100	4	0100
5	101	5	0101
6	110	6	0110
7	111	7	0111
8	1000	10	1000
9	1001	11	1001
10	1010	12	0001 0000
11	1011	13	0001 0001
12	1100	14	0001 0010
13	1101	15	0001 0011
14	1110	16	0001 0100
15	1111	17	0001 0101
16	10000	20	0001 0110
17	10001	21	0001 0111
18	10010	22	0001 1000
19	10011	23	0001 1001

Deux autres systèmes utilisés : l'octal (8 étant une puissance de 2, la conversion binaire-octal est immédiate par simple groupement des ubis trois par trois) et le décimal codé binaire (chaque chiffre décimal étant représenté en binaire). Ce dernier système facilite la conversion binaire-décimal, lors de l'entrée et des sorties, la machine travaillant nécessairement en binaire et l'utilisateur aimant lire en décimal. Ce système consomme 4 *n* ubis au lieu de 3,3 *n* pour l'écriture d'un nombre décimal de *n* chiffres : il est redondant.

comment la machine s'y reconnaît-elle?

Résumons ce que nous avons appris : *il y a toute une hiérarchie de mémoires. D'abord la cellule élémentaire, susceptible de prendre deux états; en groupant N de ces unités, on obtient un organe susceptible de 2^N états, c'est-à-dire susceptible de représenter tout nombre entier compris entre 0 et $2^N — 1$, ce nombre se trouvant codé en binaire à partir du moment où l'on a convenu qu'un état correspondait à 0, l'autre à 1. En groupant ces unités 6 par 6 (7 par 7 si l'on exige un contrôle de parité), on peut représenter un caractère alphanumérique (64 états possibles). On peut à leur tour grouper chiffres, nombres et caractères alphanumériques en mots.*

La machine saura-t-elle distinguer un chiffre d'un caractère? Saura-t-elle reconnaître le signe d'un nombre et la place de la virgule?

Généralement et normalement, les unités des mémoires sont groupées pour former des *mots* de longueur constante (d'autres modes de fonctionnement sont possibles, les machines de gestion et celles d'usage proprement logique pouvant être amenées à considérer les unités des mémoires une à une; sur certains modèles la longueur du mot est variable, et l'utilisateur peut la changer en cours de calcul). Quoi qu'il en soit, au moment précis où une instruction est lue et va être exécutée, l'unité centrale qui organise les calculs connaît la longueur du mot (par exemple : qu'il s'étend de la cellule 109 à la cellule 140).

Un calculateur à usage scientifique travaille le plus souvent avec 10 chiffres décimaux significatifs (le nombre d'opérations arithmétiques à effectuer est en général tel qu'une si grande précision n'est que rarement superflue, et il n'est pas rare qu'il faille travailler avec une beaucoup plus grande précision). Ceci correspond à 31 caractères binaires ($2^{31} = 2\,147\,483\,648$), auxquels il faut bien ajouter l'information « signe », qui coûte précisément une unité de mémoire (puisqu'il n'y a que deux signes, chacun représentable par un des deux états possibles de la cellule correspondante). On est conduit à grouper les unités en mots de 32 cellules (ce qui est commode, 32 étant une puissance de 2) ou de 33 (si l'on veut un contrôle de parité).

On a vu incidemment comment était résolue la question du signe. Mais comment le sera celle de la position de la virgule dans la représentation d'un nombre décimal ou binaire?

Deux modes d'écriture sont possibles, et chacun d'eux a ses avantages et ses inconvénients. Nous les exposerons sur les chiffres décimaux pour ne pas dérouter le lecteur, les mêmes techniques étant valables et employées par la machine en binaire.

a) La virgule fixe. On l'emploie toutes les fois que l'on connaît l'ordre de grandeur des nombres manipulés. Si un grossiste établit des factures, il sait que les prix unitaires sont des multiples de 1 centime, et que le total sera compris entre 1 et, disons, 100 000 000 centimes. Il prendra donc comme unité le centime et ne traitera que des nombres entiers susceptibles d'être lus (puisque inférieurs à 2^{31}). Ce n'est que lors de l'entrée ou de la sortie des résultats qu'il enlèvera ou mettra une virgule après le deuxième chiffre à partir de la droite, de manière à s'exprimer en unités légales : en francs et non en centimes. Cette suppression, puis cette adjonction de la virgule à une place fixe est affaire de programmation.

b) La virgule flottante. Dans les calculs scientifiques, l'ordre de grandeur des résultats physiques est également connu au moins approximativement, mais les phases de calcul purement mathématique peuvent faire intervenir des nombres considérablement plus grands ou plus petits, et il arrive que la soustraction de deux nombres très grands (de l'ordre de dix milliards, par exemple) donne un résultat très faible (de l'ordre de quelques dizaines, par exemple). L'imprévisibilité de telles opérations rend particulièrement délicate l'affectation de la position des unités à un nombre représenté dans un mot et risque d'entraîner des pertes d'information.

On indique la position de la virgule tout simplement à côté. Le nombre 513,27 sera représenté par 3·51327 (ce qui signifie que la partie entière possède 3 chiffres significatifs et se lit $0{,}51327 \cdot 10^3$), et 5132,7 par 4·51327.

Supposons par exemple que l'on ait laissé 2 positions décimales (ou 7 positions binaires) pour l'exposant et 10-2 = 8 pour l'écriture du nombre (mantisse). Le plus petit nombre sera 00·10000000 ($0{,}1 \times 10^0 = 0{,}1$), et le plus grand 99·99999999 ($0{,}99999999 \times 10^{99} \approx 10^{99}$), les nombres pouvant varier dans un facteur 10^{100}, au lieu de 10^{10}.

Signalons une dernière possibilité : si 10 chiffres significatifs (ou 8 chiffres « en flottant ») ne suffisent pas, on peut travailler en *« double longueur »* : chaque mot occupe alors deux mémoires, ce qui signifie que l'on travaille avec 20 chiffres significatifs (20 — 2 = 18 « en flottant »). Les opérations deviennent plus longues et plus encombrantes, et il faut les programmer spécialement si des instructions particulières d'arithmétique à 20 chiffres n'ont pas été prévues par le constructeur.

lettre ou symbole mathématique?

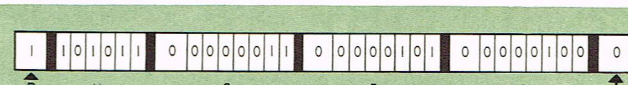
Une dernière question : une mémoire pouvant contenir un nombre ou des symboles alphabétiques (nom d'un client, lettre M signifiant qu'il y a une multiplication à faire, etc.), comment la machine s'y retrouve-t-elle?

Nous verrons que l'ordre de multiplier le contenu d'une mémoire U par celui d'une mémoire V et d'envoyer le résultat en W se code M. U. V. W.

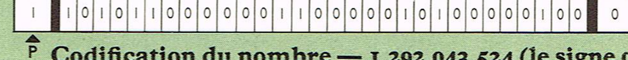
Il faut 6 positions binaires pour représenter une lettre, 8 pour chacun des numéros des mémoires (si la machine possède $2^8 = 256$ mémoires, ce qui est le cas dans la machine considérée ici), 1 pour le contrôle de parité; soit 31 unités. L'instruction peut tenir dans un mot; il restera même une position inoccupée, puisqu'une mémoire est supposée formée de 32 éléments.

Finalement, le contenu de la mémoire peut correspondre à deux significations totalement différentes.

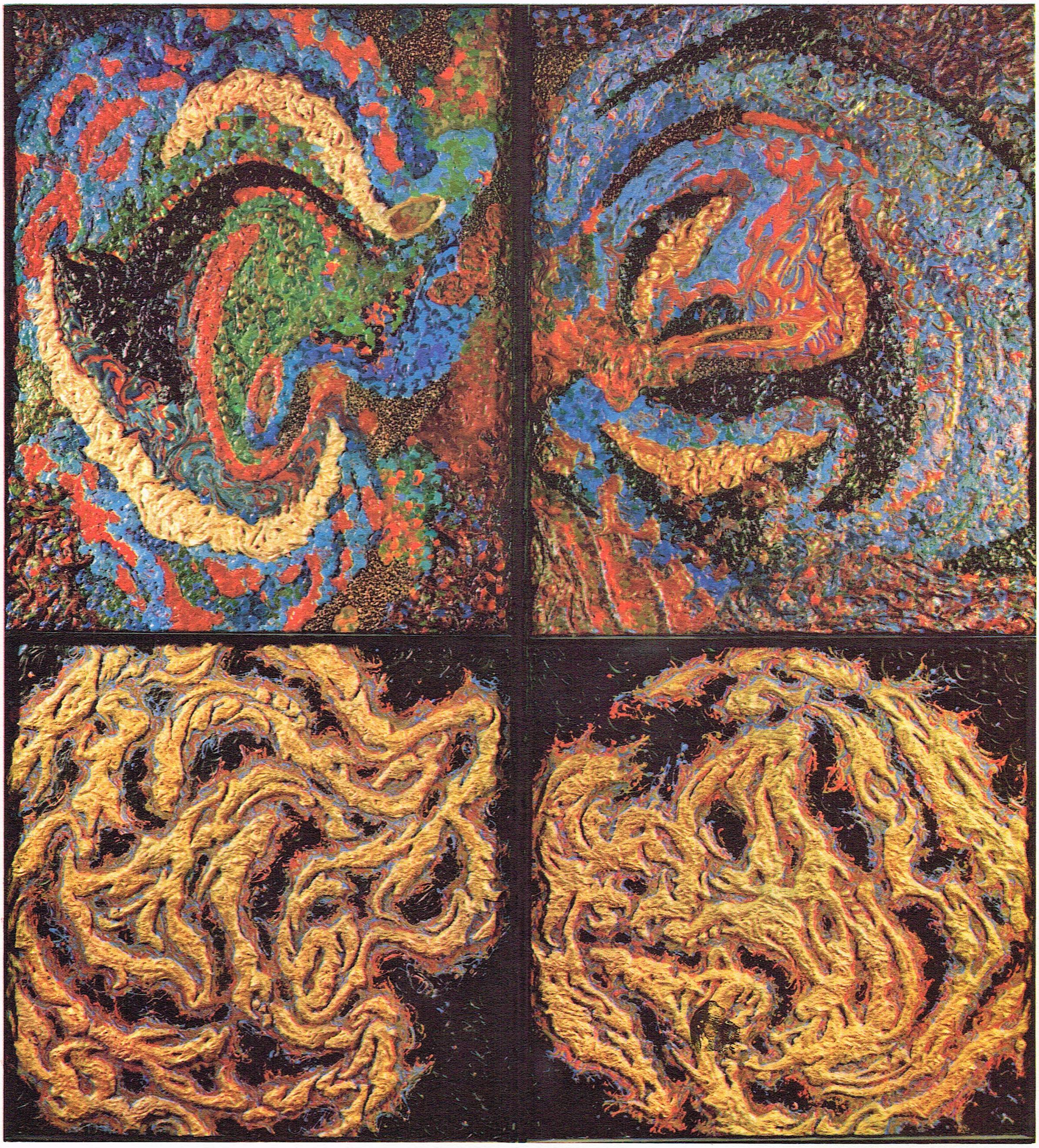
Eh bien, la machine ne les reconnaît pas directement. Si elle a reçu l'ordre de multiplier le contenu de la mémoire A, elle lit le mot comme s'il était un nombre. Si, au contraire, elle doit chercher l'instruction suivante dans la mémoire A, elle « sait » que le contenu de A est un ordre et qu'il doit donc être interprété selon le premier mode; ses systèmes logiques se prépareront donc à lire le mot A (la 32ᵉ position ne sera pas prise en considération) et à effectuer la multiplication indiquée (du même coup elle interprétera les contenus de U et de V comme des nombres). Cette possibilité de considérer le contenu d'une mémoire comme une instruction, lorsque les signaux sont dirigés vers les circuits de contrôle des ordres, ou comme une donnée, lorsqu'ils sont transférés vers un registre de calcul, est utilisée pour la modification des instructions. Le souci d'éco-



Ordre donné à la machine : « multiplier le contenu des mémoires 3 et 5 et envoyer le résultat en 4 » : C (4) = C (3) × C (5). La trente-deuxième position est arbitraire; il y a un 1 en P (de manière à avoir un nombre pair de 1).



Codification du nombre — 1 292 043 524 (le signe 0 en 32ᵉ position signifie « nombre négatif »).



PREMIÈRE MANIFESTATION DE LA PEINTURE ALGORITHMIQUE

Deux toiles de Lauquin exposées à Paris en février 1963. Un schéma de Nicole Picard, faisant appel, d'une part, à des chapitres avancés des mathématiques, d'autre part à une application numérique sur ordinateur, a fourni la structure de ces œuvres. Les parties supérieures de ces deux toiles sont la traduction fidèle de ces structures algorithmiques. les parties inférieures en sont une libre interprétation.

nomie des mémoires conduit à réutiliser plusieurs fois, au cours d'un même calcul, certaines séquences très analogues ne différant que par quelques ordres. Les contenus des mémoires correspondant à ces ordres seront transférés à titre de données en vue de subir une modification (par exemple instruction 32 au lieu de 43 pour commander une addition au lieu d'une multiplication; ou instruction 108 au lieu de 107 pour opérer sur la mémoire 108 après avoir opéré sur la précédente), puis reprendront place au mot de départ, où leur prochaine utilisation sera faite au titre d'instruction.

L'UNITE CENTRALE A LA FOIS DIRECTEUR, INGÉNIEUR ET EXÉCUTANT

Il est bon de savoir stocker l'information et de le faire de manière à la fois compacte (en utilisant au mieux la gamme des mémoires à sa disposition) et efficace (c'est-à-dire de manière que l'accès en soit rapide).

Encore faut-il opérer en fonction d'un programme préalablement établi, comportant des opérations arithmétiques, des décisions logiques et des transferts de mémoire à mémoire. Nous verrons même que l'ordinateur, méritant ainsi son nom, effectue le plus volontiers et le plus souvent de telles opérations.

Bien entendu, il faut également dire à la machine ce qu'elle doit faire. Pour cela, on code la liste des opérations à effectuer, et on l'enregistre dans les mémoires à l'instar des autres données. *Il est courant en effet qu'un programme soit décomposé en plusieurs milliers d'instructions, et il ne peut être question de prévoir un système spécial de stockage pour elles. C'est même cette possibilité de « banaliser » ordres et données qui constitue la trouvaille de génie de von Neumann, permettant d'envisager des machines à traitement automatique de l'information. C'est elle aussi qui rend possibles les modifications d'instruction que nous venons de voir.*

C'est l'unité centrale, véritable organe de direction de l'ordinateur, qui a pour fonction de rassembler et de traiter les informations au moyen des instructions successives du programme, dont elle gère le déroulement en liaison avec les autres organes, assumant ainsi un ensemble extrêmement complexe de fonctions, que nous allons examiner successivement.

1+1=1, logique propre aux ordinateurs !

A l'instar des mémoires, l'unité centrale travaille en binaire, plus exactement elle est formée de cellules interconnectées dans un réseau très dense, qui sont pratiquement toutes de l'un ou l'autre des trois types suivants : OU, ET, NON.

On peut être surpris de voir apparaître ces notions logiques là où l'on s'attendrait à trouver du numérique. La raison en est très simple : le binaire est un système spécialement adapté aux raisonnements et calculs logiques. Pour s'en assurer, il suffit d'appeler un état possible *vrai* et l'autre *faux* (chargé ou non chargé; magnétisé, non magnétisé; ouvert, fermé;...).

C'est au grand logicien anglais Boole que revient le mérite d'avoir étudié une algèbre — désignée par son nom —, qui est celle même qu'utilisent les machines électroniques. Il y a un siècle une telle algèbre n'avait aucune application et semblait être un simple jeu de l'esprit. Ainsi les travaux de Boole sont-ils tombés dans l'oubli jusqu'au moment où l'on s'aperçut que c'était sur eux (ainsi que sur la découverte de von Neumann, citée au paragraphe précédent; et bien entendu sur les progrès technologiques réalisés en ce qui concerne les composants électroniques) qu'il fallait faire reposer une industrie dont le chiffre d'affaires atteint des centaines de milliards de francs.

Nous retrouvons ici incidemment la liaison étroite qui existe entre sciences et techniques, et nous constatons le fait qu'on ne peut jamais dire que telle découverte scientifique est plus utile ou plus inutile qu'une autre, l'avenir rendant souvent justice des affirmations centrées sur des applications immédiates.

Finalement entre deux variables a et b, on peut définir les trois opérateurs suivants :

l' « union », que l'on note U (ou +)

$a \cup b$ est vrai si a **ou** b est vrai;

l' « intersection », que l'on note $a \cap b$ (ou .)

$a \cap b$ est vrai si a **et** b sont vrais;

l' « exclusion » (a et b ne peuvent pas être vrais à la fois).

On introduit aussi la négation de a, que l'on note $\bar{a}$.

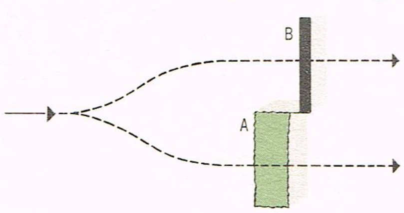
On peut s'amuser à vérifier quelques propriétés du genre « la négation de l'intersection est égale à l'union des négations » :

$$\overline{a \cap b} = \bar{a} \cup \bar{b}.$$

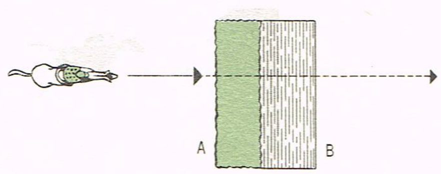
comment une machine calcule-t-elle donc ?

Ainsi donc chaque organe élémentaire n'est capable que de réaliser une des opérations de l'algèbre de Boole ou de réaliser des transferts d'information. Chacune de ces opérations ne porte que sur des éléments susceptibles de prendre deux états magnétiques ou électriques. Le passage d'un état à un autre est extrêmement rapide, de l'ordre du dix-millionième de seconde dans les machines les plus modernes; si rapide même que l'on arrive désormais à être limité par la longueur des câbles : en un temps aussi court l'impulsion électronique ne peut parcourir que quelques dizaines de mètres, et si deux

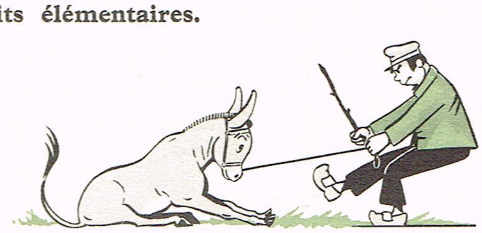
Une analogie animale des circuits élémentaires.



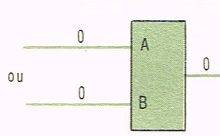
Un obstacle qu'un cheval peut franchir par l'un ou l'autre des deux parcours.



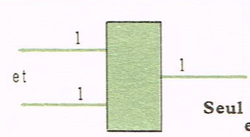
Une haie avec rivière : son franchissement nécessite de sauter A et B.



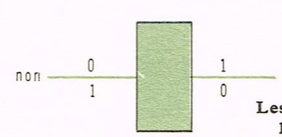
Un âne rétif.



On a représenté le seul cas où la sortie est à l'état o.



Seul cas où la sortie est excitée.



Les deux états possibles.

organes se trouvent séparés de 1 ou 2 m, le temps de transfert de l'impulsion de l'un à l'autre n'est plus négligeable; on est amené à l'idée de machines de dimensions extrêmement réduites, ne dépassant pas un mètre. Il importe également que la consommation d'énergie y soit réduite, sinon il se produirait un échauffement excessif, et il n'est pas étonnant qu'on envisage des machines fonctionnant à l'aide de supraconducteurs. L'apparition des transistors, puis de circuits intégrés très compacts et la miniaturisation des tores de ferrite ont déjà permis d'atteindre le nanoseconde.

Fermons cette parenthèse pour revenir au sujet qui nous occupe : le fonctionnement de ces machines
L'addition se traduit simplement :

$$0 + 0 = 0$$
$$0 + 1 = 1$$
$$1 + 0 = 1$$
$$1 + 1 = (1)0$$

mais un schéma booléen assez complexe est déjà nécessaire pour réaliser cette opération élémentaire, par suite notamment de la possibilité d'un report dans le quatrième cas. Nous laissons au lecteur curieux le soin de vérifier que le schéma indiqué ci-contre est correct.

Les règles de la multiplication élémentaire sont beaucoup plus simples :

$$0.0 = 0$$
$$0.1 = 0$$
$$1.0 = 0$$
$$1.1 = 1.$$

Un circuit « et » suffit à le réaliser : pour que la sortie soit dans l'état 1, il faut que l'entrée *a* **et** l'entrée *b* soient dans l'état 1.

Il ne nous est pas possible de montrer ici comment une multiplication complète s'opère en binaire, comme en arithmétique décimale, à coups de décalages et d'additions des sommes partielles.

Il se pose du reste des problèmes extrêmement intéressants, ceux dans lesquels il s'agit de savoir quel circuit réalise l'opération la plus efficace, décalages, reports et additions s'effectuant simultanément. De très belles astuces ont été trouvées et aboutissent à des circuits de multiplication de plus en plus efficaces.

On n'est pas surpris de voir que l'organe chargé des quatre opérations élémentaires (addition, multiplication, soustraction et division) est extrêmement complexe et peut grouper des dizaines de milliers d'éléments, dont chacun a une fonction bien particulière, interconnectés suivant un système de câblages ou de circuits d'une complication étonnante. A plus forte raison si la machine possède le câblage nécessaire pour faire les opérations en double longueur ou à virgule flottante.

Nous avons vu que chaque mot possédait un nombre de chiffres binaires p bien déterminé, par exemple $p = 32$.

La somme de deux de ces mots ne peut avoir plus de $p + 1$ chiffres, et encore ce dernier cas ne se produit que s'il y a un report sur les unités de gauche, par exemple comme lorsqu'on écrit en calcul décimal $7 + 8 = 15$. De même, le produit de tels nombres ne peut avoir plus de $2p$ chiffres, de sorte que tous les résultats puissent tenir dans deux mémoires. En cas de division, on stocke le quotient pour lequel il suffit de considérer p chiffres significatifs dans l'une de ces deux mémoires, et le reste dans l'autre.

Ainsi les opérations ne portent que sur deux mémoires, et il importe de transférer d'abord les deux facteurs, puis, après l'opération, d'envoyer le résultat à l'endroit désiré, tout en libérant la mémoire-opérateur.

Nous commençons à entrevoir qu'un ordinateur passe une bonne partie de son temps à opérer des choix logiques ou des transferts de nombres.

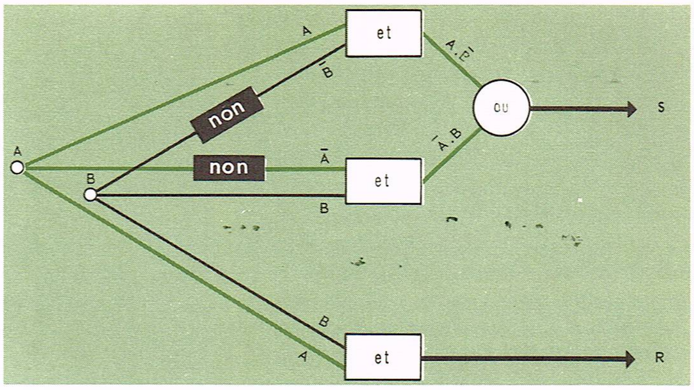


Schéma de l'additionneur à une position binaire.
2 entrées : A et B à valeur 0 ou 1.
2 sorties : S, somme de A et B : position binaire origine ;
R report de A + B sur la position binaire voisine.
Il y a quatre cas possibles analysés dans le tableau ci-dessous, où sont notées :
— les valeurs à l'entrée ;
— la formation de S, qui est la somme exclusive de A et B (S vaut 1 si soit A, soit B, et l'un des deux seulement, vaut 1) ;
— la formation du report par le seul connecteur *et*.
Il est nécessaire d'utiliser des éléments logiques formant le *et*, le *ou*, le *non*.

A	B	formation de S				S	R
		$\bar{A}$	$\bar{B}$	$A.\bar{B}$	$\bar{A}.B$		A.B
0	0	1	1	0	0	0	0
0	1	1	0	0	1	1	0
1	0	0	1	1	0	1	0
1	1	0	0	0	0	0	1

Remarquons une dernière fois que la machine n'opère généralement jamais sur des nombres, mais sur des contenus de mémoires. M.3, 4 veut dire multiplier le contenu de la mémoire 3 par celui de la 4, et non 3 par 4. Si l'on veut multiplier 3 par 4, il faut lire 3, le ranger dans la mémoire m_1, de même pour 4 qui sera rangé dans m_2, et donner l'ordre M, m_1, m_2.

une curieuse arithmétique pour des opérations nécessairement fausses

En fait, il n'importe pas de conserver tous les chiffres du produit; il faut arrondir, en éliminant purement et simplement un certain nombre de chiffres à droite, dépourvus de valeur. Le calculateur manuel estime à chaque opération ce qu'il faut conserver pour maintenir la précision à un niveau raisonnable. Il écrit, par exemple,

$$1{,}761 . 3{,}18 + 2{,}513 = 8{,}1130 \text{ (et non } 8{,}11298).$$

L'ordinateur va opérer différemment, et beaucoup plus systématiquement : il éliminera les p chiffres de droite, de manière à faire tenir chaque résultat, comme chaque donnée, dans une mémoire exactement. Par exemple, avec 4 chiffres significatifs, on aura $1{,}761 . 03{,}18 + 2{,}513 = 8{,}113$.

Quoi qu'il en soit, la limitation à droite des chiffres représentés dans une mémoire est une source d'erreurs, que nous allons

mettre en évidence pour un cas élémentaire. Soit à calculer $\frac{7 \times 5}{3}$ avec 4 décimales; en effectuant d'abord la division, on trouve $\left(\frac{7}{3} = 2{,}333\,3\right) \cdot 5 = 11{,}666\,5$.

Si l'on commence par la multiplication, il vient
(7.5 = 35) : 3 = 11,666 6 (ou 11,666 7 après avoir arrondi sur le dernier chiffre).

Le fait de travailler sur quatre décimales détermine une erreur sur le dernier chiffre, et les résultats sont alors différents. Le programme devra tenir compte de ce principe élémentaire contraire à l'arithmétique la plus classique : l'ordre des opérations intervient !

comment calcule une machine... qui ne sait même pas l'algèbre ?

Comment a donc pu opérer cet ordinateur « Univac » qui vient de résoudre — en seize secondes — un système de cent équations à cent inconnues?

Vous avez appris que la solution du système :

$$\begin{cases} a_1x + b_1y = c_1 \\ a_2x + b_2y = c_2 \end{cases} \text{ est } x = \frac{c_1b_2 - c_2b_1}{a_1b_2 - a_2b_1},\ y = \frac{a_1c_2 - a_2c_1}{a_1b_2 - a_2b_1}$$

(si le dénominateur commun n'est pas nul), Pour résoudre le système $\begin{cases} 4x + 5y = 17 \\ x + 8y = 12 \end{cases}$ il suffit de remplacer les lettres par la valeur numérique qu'elles représentent

$$x = \frac{17.8 - 12.5}{4.8 - 1.5} = \frac{76}{27} = \ldots \quad y = \frac{4.12 - 1.17}{4.8 - 1.5} = \frac{31}{27} = \ldots$$

La machine procédera-t-elle de manière analogue, en cherchant l'expression algébrique du résultat (quelle qu'en soit la complexité pour un système de cent équations à cent inconnues) et en appliquant la formule trouvée au cas particulier demandé?

Là encore la réponse est « non »; la machine ne connait pas l'algèbre littérale, et ne peut — du moins dans l'état actuel de la technique et de la programmation — effectuer que des opérations arithmétiques, logiques et des transferts portant sur des caractères représentés dans un système convenable de numération, et non sur des symboles algébriques. Autrement dit encore, on opère sur des mots contenus dans des mémoires dûment numérotées, mais non sur les mémoires elles-mêmes.

Reprenons la résolution du système de deux équations à deux inconnues, et examinons en détail le processus du calcul automatique qui le permet. Le calcul du dénominateur $a_1b_2 - a_2b_1$ exige deux multiplications et une soustraction. Supposons que 101, 102, 103, 104 soient les *adresses* (ou numéros de mémoires) où sont rangés les coefficients au début du calcul. Il faudra :

— effectuer a_2b_1 (opération arithmétique);

— ranger ce produit dans une mémoire de transit, 107 par exemple (c'est l'opération de transfert);

— effectuer a_1b_2;

— effectuer la soustraction $a_1b_2 - a_2b_1$;

— vérifier si $a_1b_2 - a_2b_1 \neq 0$ (opération test);

— transférer le résultat vers la mémoire 107 réutilisable, pour conserver la valeur calculée $a_1b_2 - a_2b_1$;

— effectuer des calculs analogues au moyen des mémoires 108 et 109 pour obtenir les valeurs des numérateurs des formules qui donnent x et y;

— effectuer les deux divisions pour obtenir x et y;

— si la première opération test a montré que $a_1b_2 - a_2b_1 = 0$, on est conduit à d'autres opérations tests : $c_1b_2 - c_2b_1$ et $a_1c_2 - a_2c_1$ sont-ils nuls ou non?

— si oui, il y a une infinité de couples (x, y) pour résoudre le système posé en exemple : le programmeur sera averti par l'impression d'un code spécial (opération de *sortie*);

— si non, il n'existe pas de solution, ce qui devra être signalé, de même, par l'impression d'un autre code choisi d'avance.

Les éléments de l'unité centrale n'ont eu à manipuler, au cours de cette résolution, ni x, ni y, ni aucune formule toute prête, mais seulement des quantités logées dans différentes cases et sur lesquelles les opérations ont été des multiplications, des soustractions, des divisions, des tests et des sorties. De plus, la séquence d'instructions suivie (repérée à l'aide de deux flèches) aura nécessité au moins une opération de *saut*.

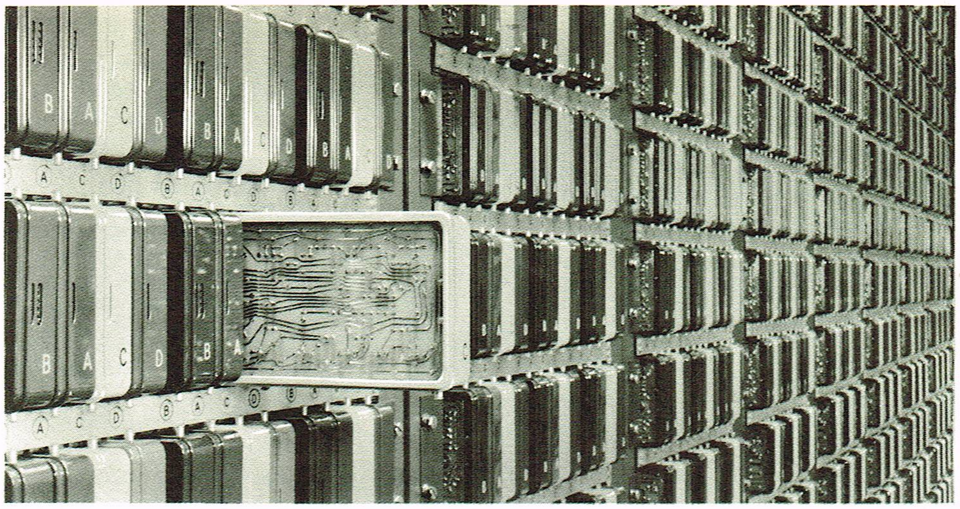
Quel avantage avons-nous gagné en remplaçant les opérations arithmétiques, que l'on peut effectuer en cinq ou dix minutes, par l'écriture — longue et délicate — d'un programme (une heure de travail au moins) qu'il nous faudra passer dans une machine dont une minute de fonctionnement peut coûter plusieurs centaines de francs?

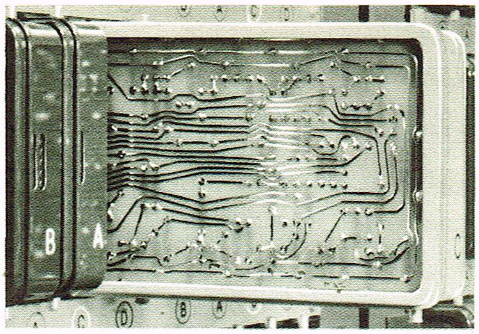
Aucun, si le calcul ne porte que sur un système unique; un grand, s'il y a un nombre suffisant de problèmes analogues à résoudre, ou si le système considéré porte sur cent équations à cent inconnues par exemple. S'il y avait eu 1 000 cas à résoudre, il aurait fallu 1 000 fois 10 minutes pour les faire à la main, mais toujours une heure pour écrire le programme, qui est le même pour un système ou pour des milliers. **Le calcul électronique finit toujours par devenir rentable lorsqu'il porte sur un nombre suffisant de problèmes analogues, ou sur un grand problème décomposable en petits problèmes analogues (même formule à employer, portant sur des données différentes), c'est-à-dire lorsqu'il est répétitif.**

Si l'utilisateur désireux de résoudre un système de deux équations du premier degré à deux inconnues écrit un programme analogue à celui que nous avons proposé, celui qui devra considérer des systèmes de trois équations ne pourra s'en servir : il lui faudra écrire un autre programme pour « réaliser » des formules différentes, et un autre pour quatre équations, etc. Il y aura un programme différent pour chaque ordre. De plus, les formules algébriques deviennent inextricables, tant elles comportent de termes ($n!$ pour l'ordre n, c'est-à-dire environ 4 000 000 pour l'ordre 10). Ce fut le rôle des théoriciens du calcul matriciel que d'inventer des procédés **(algorithmes)** pour résoudre numériquement tout système d'équations du premier degré, de quelque ordre qu'il fût. (La méthode la plus classique, la plus célèbre, la plus simple, et probablement la meilleure, connue sous le nom de « méthode de Gauss », n'est en fait qu'une transformation mathématique de la méthode d'élimination des inconnues une par une, telle qu'on l'enseigne dans les lycées.) Il suffit de faire entrer dans la machine les coefficients de l'équation, son ordre et le programme valable quel que soit celui-ci (il faut que la machine ait la place nécessaire au stockage des données et des résultats intermédiaires) pour que se déclenche l'algorithme de calcul et que s'imprime la solution (ou l'une des indications « système impossible », « système indéterminé », ou ... « machine en panne »).

pourquoi des machines si puissantes ?

Quels sont le coût et le bénéfice de l'opération? Depuis longtemps, cet algorithme de Gauss a été programmé sur toutes les machines, et il a été utilisé des millions de fois. On peut donc admettre qu'un tel programme — livré avec la machine par le constructeur, qui a pour tâche essentielle de vendre non seulement une machine, mais encore une « bibliothèque » de programmes élémentaires, évitant à chaque utilisateur de refaire ce qui a déjà été rédigé plusieurs fois — ne coûte que





« Calculateur » de la centrale nucléaire de Chinon. Un panneau ouvert montre un circuit imprimé, contenant également une dizaine de transistors, tel qu'il en existe quelques dizaines de milliers — d'environ 30 modèles différents — dans une machine très puissante.

le temps nécessaire pour le faire entrer dans la machine. Reste le coût de la lecture des données et de l'impression des résultats, et le coût du calcul proprement dit. La lecture est toujours lente; son coût C_1 est proportionnel au nombre de données, c'est-à-dire au carré n^2 de l'ordre (n lignes et n colonnes) : $C_1 = K_1 n^2$, où K_1 est très grand.

Le coût C_2 du calcul varie, lui, comme le nombre d'opérations, c'est-à-dire comme n^3 (dans la méthode d'élimination et dans presque toutes les méthodes connues) : $C_2 = K_2 n^3$.

K_2 étant beaucoup plus petit que K_1 (le calcul est beaucoup plus rapide que l'entrée ou la sortie, le facteur pouvant atteindre 1 000), C_2, très petit lorsque n est faible (la machine passe tout son temps à lire et à imprimer, et non à calculer), croît beaucoup plus vite que C_1. Il faut que n soit suffisamment grand pour que la machine se comporte vraiment comme un organe de calcul, et non de lecture, et devienne efficace.

Pratiquement, le calcul manuel ne permet pas de dépasser $n = 10$: un système de dix équations à dix inconnues représenterait plusieurs semaines de travail. Alors qu'il n'est pas rare de voir des calculateurs attelés à la résolution de systèmes d'ordre 100, 1 000 (le travail manuel exigerait plusieurs millions de semaines!), voire même 10 000, pour des systèmes particulièrement simples il est vrai (problèmes de recherche opérationnelle ou de transport).

On retrouve la conclusion précédente : les machines sont irremplaçables lorsque les calculs sont très répétitifs, ou très longs, et qu'ils portent sur des problèmes très difficiles. Elles permettent d'aborder des problèmes — théoriques ou pratiques — impossibles à envisager par le calcul manuel. Ce qui conduit à construire des machines de plus en plus rapides et de plus en plus puissantes (à grande mémoire).

un ordinateur : une machine qui n'est pas seulement occupée à calculer

Ainsi donc même les calculs arithmétiques élémentaires impliquent un certain nombre d'opérations telles que des décalages de longueurs variables et des transferts d'information. Pour des raisons technologiques, les instructions correspondantes ne peuvent porter que sur deux mémoires (A et Q). Le transit d'une mémoire M_1 à une mémoire M_2 se décompose en deux temps : $M_1 \rightarrow A$ (ou Q) $\rightarrow M_2$. Les transferts peuvent s'effectuer ubi par ubi (ce qui est utile lorsque la machine fait des comparaisons ou des calculs purement logiques), caractère par caractère (en utilisation alphanumérique), mot par mot ou même bloc par bloc d'informations, selon la hiérarchie des mémoires; des groupes de 64 ou 100 mots, par exemple, peuvent être « appelés » d'une unité de mémoire (disque, bande ou tambour) vers la mémoire principale de ferrite. Mais d'autres opérations sont tout autant utiles; comparaisons et tests sont des opérations capitales; on les place en des « nœuds » du programme toutes les fois que des opérations ultérieures dépendent de résultats intermédiaires. Avant de prendre la racine carrée d'un nombre, par exemple, il importe de savoir s'il est positif, nul ou négatif. Dans le premier cas, le calcul sera possible, et une première suite d'opérations s'enchaînera naturellement; dans le dernier, il faudra décider d'une autre suite d'opérations : abandon des calculs et impression d'un code signifiant que l'extraction n'a pu avoir lieu, calcul en nombres complexes, etc.

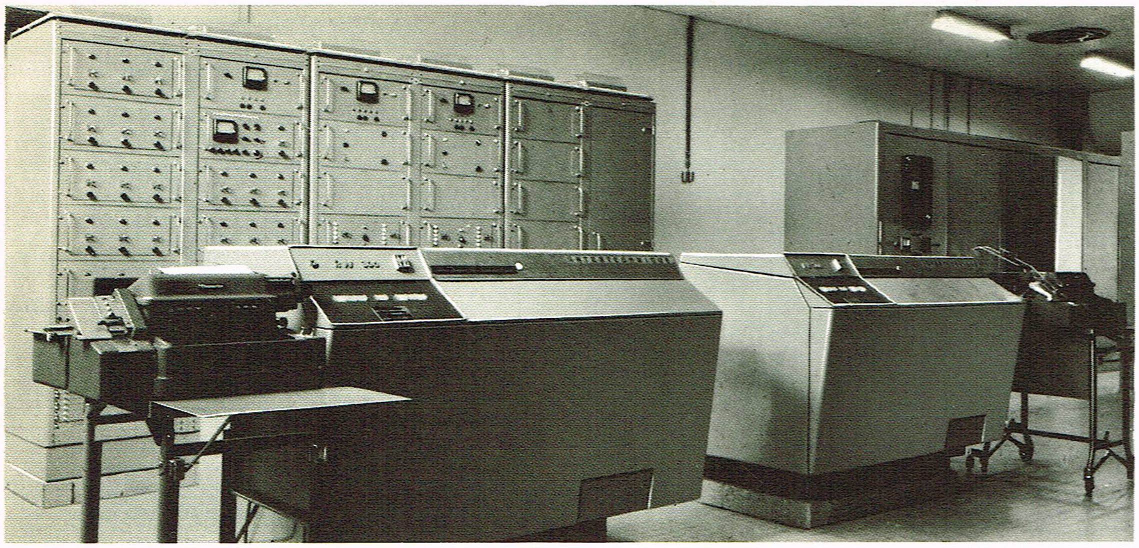
Pour en terminer avec l'unité centrale, il faut savoir que c'est elle aussi qui est chargée de dépouiller les instructions. Celles-ci figurent dans une mémoire spéciale : le registre d'instructions. Dès son arrivée, l'instruction est décodée. Dans le cas déjà indiqué, la machine lira d'abord le premier ubi (contrôle de parité) pour vérifier qu'il n'y a pas eu d'erreur lors de l'enregistrement. Elle trouvera les 6 ubis suivants et en déduira qu'elle doit se mettre en état de multiplication, c'est-à-dire ouvrir les circuits qui correspondent à cette opération. En lisant les 24 signaux suivants, elle verra qu'il faut aller chercher les contenus des mémoires U et V, et s'apprêter à envoyer le résultat dans la mémoire W. On voit quel ensemble important de circuits elle doit préparer avant que commence l'opération proprement dite. De plus, dès que le registre d'instructions sera libéré (au début de la multiplication proprement dite), elle lira l'instruction suivante et préparera sa mise en place. Un registre spécial, appelé « compteur ordinal », comptera le nombre d'instructions exécutées, de manière à savoir à chaque instant où en est l'exécution du programme.

Il y a encore d'autres organes constituant eux aussi l'unité centrale. A côté des mémoires opératrices du registre d'instructions et du compteur ordinal, il faut essentiellement mentionner les rythmeurs, véritables métronomes chargés d'assurer le fonctionnement synchronisé de tous les organes, et signaler l'existence des registres spécialement adaptés à la modification des adresses figurant dans les opérations (registres d'index).

le service des relations extérieures

Les supports matériels de l'information qu'utilisèrent les ordinateurs classiques sont les cartes, puis les rubans, perforés à partir de documents de base ou par des moyens automatiques commandés par des instruments de mesure. Ces supports permettent l'introduction de l'information soit directement dans l'unité de mémoire, soit dans les registres de transit. De toute façon, l'unité centrale prend en charge l'information qui entre

Phot. C. S. F.-Georges Bru.



Une machine travaillant en temps réel.
Dans cette calculatrice RW 300, 128 canaux spéciaux permettent l'introduction de données prioritaires en provenance de phénomènes de nature variée pouvant être transformés en impulsions électriques. Des calculs effectués au fur et à mesure de l'arrivée de ces données permettent, par exemple, la surveillance et la conduite d'un cracking, de la production de neutrons dans un réacteur nucléaire, du trafic d'un aéroport.

pour la diriger vers l'organe de stockage désiré : mémoire principale ou bande magnétique. Les performances des machines à cartes sont passées de 150 à 300, puis à 650, et aujourd'hui à 1 200 cartes à la minute, permettant l'introduction de nombres binaires sur 80 colonnes de 12 chiffres. La rapidité des lecteurs de rubans n'a pas crû dans les mêmes proportions, mais dans les deux cas la vitesse a pu être augmentée par l'emploi de cellules photo-électriques.

Une autre entrée, moins usuelle et pourtant adaptable sur beaucoup de calculateurs, est constituée par tout organe supplémentaire qui reçoit des impulsions provenant d'un phénomène physique et les traduit sous forme de chiffres (on dit qu'elle les « digitalise »). Ce peut être un appareil de mesure ou de détection, ou un mécanisme. L'information reçue par le calculateur est alors dite « *en temps réel* », car elle arrive à la cadence de déroulement du phénomène et est traitée instantanément.

Les problèmes de sortie sont symétriques à ceux d'entrée, mais avec une fonction supplémentaire : celle de l'impression.

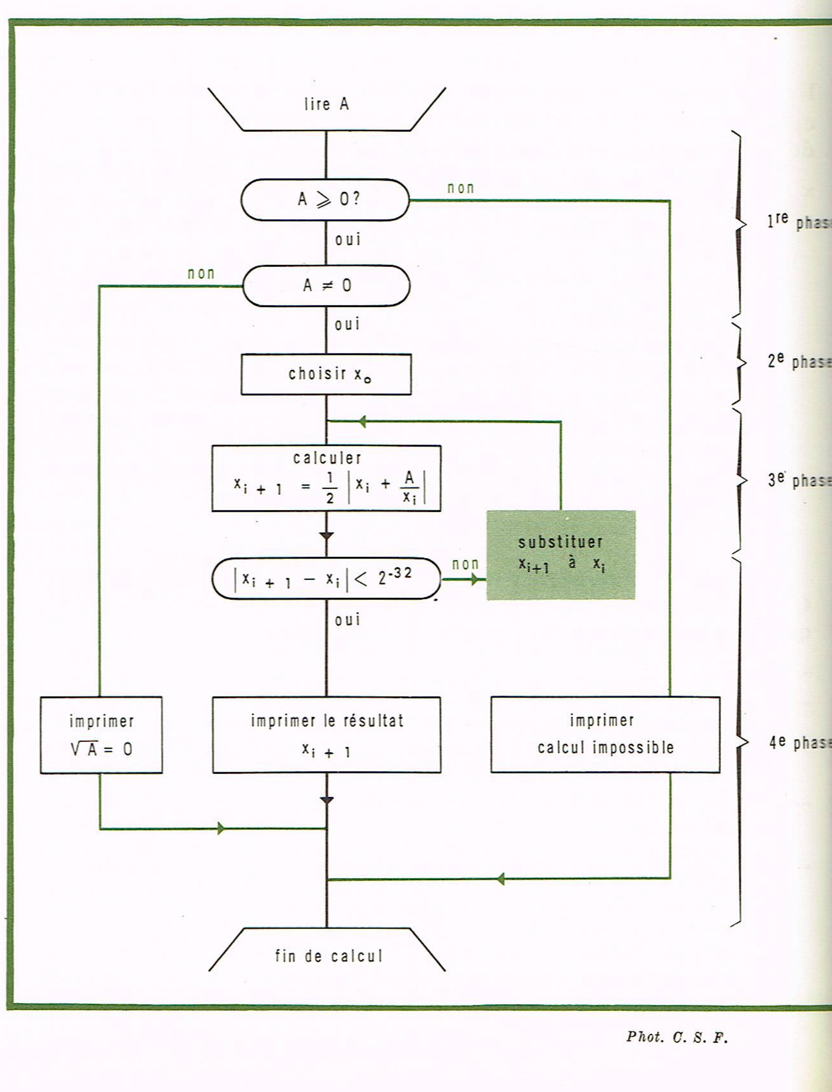
Certains calculateurs sont spécialisés dans le traitement instantané et bénéficient pour cela de lignes d'interruption, de manière à tenir compte sans délai de l'information nouvelle, qui peut arriver à travers un grand nombre de canaux physiques différents.

L'entrée et la sortie sont souvent les goulets d'étranglement du traitement de l'information, car les vitesses mécaniques ne croissent pas selon un facteur 100 en quatre ans comme les vitesses de calcul pur. C'est pourquoi il existe des calculateurs auxiliaires spécialement adaptés à ces travaux d'entrée et de sortie au moyen de supports perforés (cartes et rubans), et capables de transmettre l'information vers les bandes magnétiques du calculateur principal.

La lenteur des organes de communication de la machine avec l'extérieur est telle que l'on est amené à utiliser, dans un gros ordinateur, en une véritable organisation périphérique, de petits calculateurs préparant les données et les stockant sur une bande magnétique que l'unité centrale lira ensuite très rapidement, les opérations lentes des lectures préliminaires n'immobilisant que des éléments secondaires.

Il existe une autre conséquence de cette lenteur; elle concerne les centres mécanographiques équipés de calculateurs électroniques destinés à l'établissement de bulletins de paye ou de facturation. Les calculs (une quinzaine d'opérations pour établir et vérifier un bulletin de paye) sont si rapidement effectués que la machine est sans cesse en « attente lecture »; on peut en profiter, surtout si l'on dispose d'un bon nombre de mémoires, pour faire des calculs de gestion : inventaire de stock, état du personnel, préparation du bilan et des déclarations fiscales, etc. **Le tout est pratiquement gratuit, et le centre mécanographique révèle alors sa véritable utilité.**

Les problèmes non encore résolus sont essentiellement ceux du rassemblement des données en vue d'une perforation automatique ou de la lecture directe des documents d'entrée, et, à l'autre extrémité de la chaîne, ceux de la condensation des résultats sortis en vue de leur utilisation commune avec tous les résultats partiels « éventuellement » exploitables qui alourdissent souvent énormément le dépouillement.



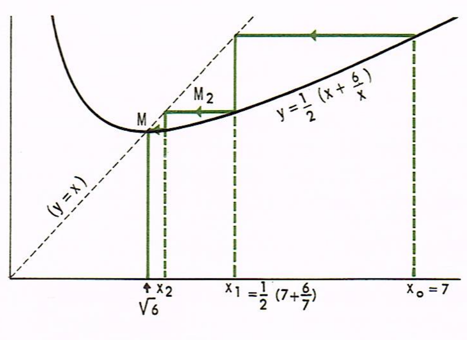
Phot. C. S. F.

LA PROGRAMMATION, ART DE SE FAIRE OBÉIR PAR UNE MACHINE

Un calculateur doit pouvoir servir à des usages variés, et ne pas se contenter de répéter indéfiniment le même cycle d'opérations, comme une machine-outil programmée une fois pour toutes et dont le fonctionnement est invariable; il faut donc lui donner des ordres. En leur absence, à la sortie de l'usine par exemple, alors que toutes ses mémoires sont vierges, le calculateur ne peut rien faire.

Un organe mécanique exécute le travail qu'on lui demande au fur et à mesure de l'arrivée des ordres; ce n'est qu'à l'instant où tourne la route que le conducteur braque le volant et que celui-ci agit sur les roues, et il n'est arrivé à aucun conducteur de calculer la trajectoire du véhicule au moment de prendre la route et de donner alors tous les ordres à l'avance pour arriver à bon port, en admettant que l'on puisse prévoir tous les incidents de la circulation. Bien au contraire, la vitesse d'exécution d'un organe électronique (de l'ordre de la microseconde) est telle, qu'il faut donner à l'avance tous les ordres, aucun homme n'étant capable de dicter un million d'instructions pendant une seule seconde. Ainsi est apparue la notion de programme, ensemble ordonné des instructions que la machine devra exécuter successivement pour résoudre le problème posé, compte tenu de toutes les circonstances et incidents susceptibles de surgir en cours de route.

Représentation géométrique de l'itération proposée pour calculer $\sqrt{A}$. On vérifiera que les points M_i se rapprochent de plus en plus vite du point M défini par $x = \frac{1}{2}\left(x + \frac{A}{x}\right)$, c'est-à-dire $x = \frac{A}{x}$ ou encore $x = \sqrt{A}$.



Calcul itératif de $\sqrt{x}$ par la méthode de Newton; on remarque quatre phases :

- **Chercher si A est positif, négatif ou nul.**
- **Initialisation : on choisit une première valeur x_0 aussi approchée que possible (il suffit, par exemple, de connaître les carrés des entiers de 1 à 10 pour avoir la valeur entière de tout nombre compris entre 0 et 100).**
- **Calcul proprement dit de nombres x_i approchant de mieux en mieux $\sqrt{A}$.**
- **Test d'arrêt, permettant de vérifier que l'approximation est suffisante (on s'arrête ici lorsque les deux valeurs successives sont égales au sens de la machine, c'est-à-dire diffèrent de moins de 2^{-32}). Bien entendu, les deux dernières phases se déroulent successivement, à autant de reprises qu'il y a d'étapes d'itérations.**

poser correctement les problèmes

La première étude d'un problème consiste à en effectuer une analyse très poussée. Il est absolument nécessaire que le problème soit bien posé, car le programmeur n'est pas responsable d'une mauvaise étude mathématique. Il revient à celui qui commande les calculs, qu'il soit organisateur ou chimiste, administrateur ou physicien expérimental, d'indiquer au mathématicien quelles sont les données dont il dispose au départ (hypothèses) et quels résultats il attend de l'intervention du calculateur (conclusion). Le mathématicien a un choix assez restreint dans la voie menant de l'une à l'autre. Se fondant sur ses connaissances du domaine numérique, il doit recommander un type de méthode de calcul avec la possibilité de faire certains choix suivant les facilités d'adaptation aux diverses machines dont dispose le centre de calcul. La comparaison détaillée des méthodes à employer permettra à l'analyste de choisir l'algorithme en tenant compte des impératifs économiques : rapidité de la programmation, vitesse de mise au point du calcul de la machine, temps d'exécution, nombre des mémoires utilisées et prix de revient.

choisir une « bonne » méthode de calcul

Trouver un algorithme (ensemble ordonné des opérations à effectuer que l'on traduit sous forme d'un programme à imposer au calculateur) susceptible de résoudre un problème numérique, mieux encore, trouver l'algorithme optimal est une branche des mathématiques qui s'appelle « analyse numérique ».

Elle se distingue de l'analyse classique, dont l'essentiel est la démonstration de théorèmes d'existence et d'unicité, en ce qu'elle doit nécessairement conduire à des procédés de calcul susceptibles, sinon de résoudre exactement le problème, du moins d'en donner la solution la plus approchée possible.

Nous ne pouvons que citer les principales branches de cette discipline récente (calcul matriciel, programmation linéaire, théories de l'interpolation et de l'intégration, résolution d'équations différentielles et d'équations aux dérivées partielles, analyse linéaire et recherche des éléments propres d'une matrice ou d'un opérateur, décomposition spectrale et analyse de Fourier) et donner les principes généraux permettant de choisir une méthode parmi toutes celles qui sont possibles.

Il est exceptionnel que l'on puisse obtenir la solution exacte d'un problème, car plusieurs raisons s'y opposent :

— Les coefficients sont nécessairement empreints d'incertitude d'origine physique (grandeur physique entachée d'erreurs) ou mathématique (représentation des nombres avec un nombre fini de chiffres significatifs);

— Chaque mémoire, donc chaque résultat, ne peut prendre qu'un nombre fini (2^N) de valeurs, ce qui entraîne l'existence d'un seuil : la machine ne peut pas distinguer un nombre rigoureusement nul d'une quantité inférieure au seuil. Ainsi donc, même les tests et les comparaisons comportent une incertitude liée aux erreurs d'arrondi;

— On ne peut faire la plupart des opérations arithmétiques élémentaires que de manière approchée : un dépassement de capacité dans l'addition ou la soustraction se traduit par la perte d'un chiffre significatif; toute multiplication se traduit par une perte de la moitié des chiffres du résultat; il en est de même pour la division, puisqu'on ne peut obtenir que les premiers chiffres d'un quotient qui, en général, en comporte un nombre infini (erreurs d'arrondi);

— Il existe rarement un algorithme comportant un nombre fini d'opérations susceptibles de résoudre le problème, et il faut définir un algorithme itératif, c'est-à-dire calculer une suite de valeurs qui se rapprochent de la solution (en termes mathématiques, « qui tendent vers la solution ») : $x_i \to x_o$.

Il faut, bien entendu, s'arrêter à un moment donné et admettre qu'une certaine valeur x_N est une approximation convenable de la solution x_o. Il est, en général, très difficile d'estimer l'erreur ainsi commise (erreur d'algorithme ou de chute). La seule chose que l'on puisse dire, c'est qu'elle n'est pas nulle.

Il faut également tenir compte du coût d'obtention de la solution, c'est-à-dire, pratiquement, du nombre d'opérations que comporte l'algorithme. **Le problème de l'analyse numérique est très simple à formuler : trouver un algorithme, généralement itératif, à la fois rapide et donnant un résultat suffisamment approché.**

cinq itérations pour calculer une racine carrée

Il importe de bien réaliser l'importance des calculs itératifs, et pour cela d'en donner un exemple facile et d'intérêt universel.

Il s'agit du calcul de la racine carrée d'un nombre positif A, 6 par exemple. Une ancienne technique, que l'on apprend encore dans l'enseignement secondaire, calcule la racine chiffre par chiffre :

$$4 < A = 6 < 9 \quad \text{donc} \quad 2 < x = \sqrt{A} < 3$$
$$5{,}76 < 6 < 6{,}25 \quad \text{donc} \quad 2{,}4 < x < 2{,}5$$
$$5{,}9536 < 6 < 6{,}0025 \quad \text{donc} \quad 2{,}44 < x < 2{,}45$$

Il s'agit d'un procédé par tâtonnement : on peut dire qu'il est itératif puisque l'on cherche deux suites de nombres (2; 2,4; 2,44; 2,449; 2,4494.....) (3; 2,5; 2,45; 2,450; 2,4495), l'une croissante, l'autre décroissante, qui approchent de mieux en mieux la solution $\sqrt{6} = 2{,}449489742...$ Mais une telle technique est longue et coûteuse : chaque itération ne donne qu'un chiffre significatif de plus, obtenu au prix de longs calculs : il faut, par exemple, calculer les carrés de 2,1, de 2,2, de 2,3, de 2,4, de 2,5, et les comparer à 6 pour être certain du premier nombre après la virgule.

On peut améliorer la méthode de la manière suivante : $2{,}45^2$ étant très proche de 6, il est vraisemblable que le chiffre des centièmes de $\sqrt{6}$ sera plus proche de 9 que de 1, d'où l'idée de calculer tout de suite $(2{,}449)^2$ et non tous les nombres $(2{,}441)^2$, $(2{,}442)^2$... En calcul manuel on peut effectivement opérer ainsi; en calcul automatique, il existe des méthodes permettant aussi de le faire.

On a été amené à élaborer des méthodes très puissantes pour résoudre une équation quelconque $f(x) = 0$. La plus célèbre est due à Newton, qui propose la formule

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

(f' étant la dérivée de f), qui revient à remplacer la courbe $f(x) = 0$ par sa tangente. Sous certaines conditions (mais non toujours), la suite des x_i converge, et converge vite.

Appliquée à l'équation $x^2 - A = 0$, dont la solution est bien $\sqrt{A}$, on aboutit à la formule $x_{i+1} = \frac{1}{2}\left[x_i + \frac{A}{x_i}\right]$.

On part d'une valeur voisine de $\sqrt{A}$ (par exemple, 3 dans l'exemple précédent), puis on calcule successivement :

$$x_1 = \frac{1}{2}\left[3 + \frac{6}{3}\right] = 2{,}5,$$
$$x_2 = \frac{1}{2}\left[2{,}5 + \frac{6}{2{,}5}\right] = \frac{2{,}5 + 2{,}4}{2} = 2{,}45,$$
$$x_3 = \frac{1}{2}\left[2{,}45 + \frac{6}{2{,}45}\right] = \frac{2{,}45 + 2{,}44898}{2} = 2{,}44949,$$
$$x_4 = \frac{1}{2}\left[2{,}44949 + \frac{6}{2{,}44949}\right] = 2{,}44948974278.$$

On voit qu'il a suffi de 4 divisions, de 4 additions et de 4 décalages (correspondant, en calcul binaire, à quatre divisions par 2) pour obtenir 12 chiffres exacts. En fait, chaque étape double le nombre de chiffres exacts : 12 au bout de 4 itérations, 24 au bout de 5, presque 200 au bout de 8 ($\sqrt{6} = 2{,}449489742783178 32...$)

On voit la puissance de cet algorithme, toujours convergent comme le montre le schéma de la page 229.

comment étudier un faisceau de trajectoires de satellite

Les calculs de trajectoires sont d'une importance primordiale pour la détermination des tirs et des lancers de satellites. Les positions successives du mobile sur sa trajectoire sont obtenues à partir des équations de la mécanique sous forme de relations faisant intervenir des dérivées par rapport au temps : $\frac{dF}{dt} = G(t)$.

$G(t)$ est une fonction du temps et de paramètres géométriques ou de masse, F est par exemple l'abscisse relativement à un certain repère, G représente la dérivée de F par rapport au temps, c'est-à-dire la pente de la tangente à la courbe $F(t)$. Or, plusieurs fonctions déduites l'une de l'autre par translation admettent la même dérivée. Il est donc nécessaire de fixer un point de la courbe à l'aide de ce qu'on appelle les constantes d'intégration, ce qui correspond à préciser les conditions au départ.

Il est quelquefois possible d'intégrer immédiatement $G(t)$, c'est-à-dire d'exprimer $F(t)$ par une formule faisant intervenir des fonctions dites essentielles, telles que sin, cos, exponentielle, log, arctg, etc.

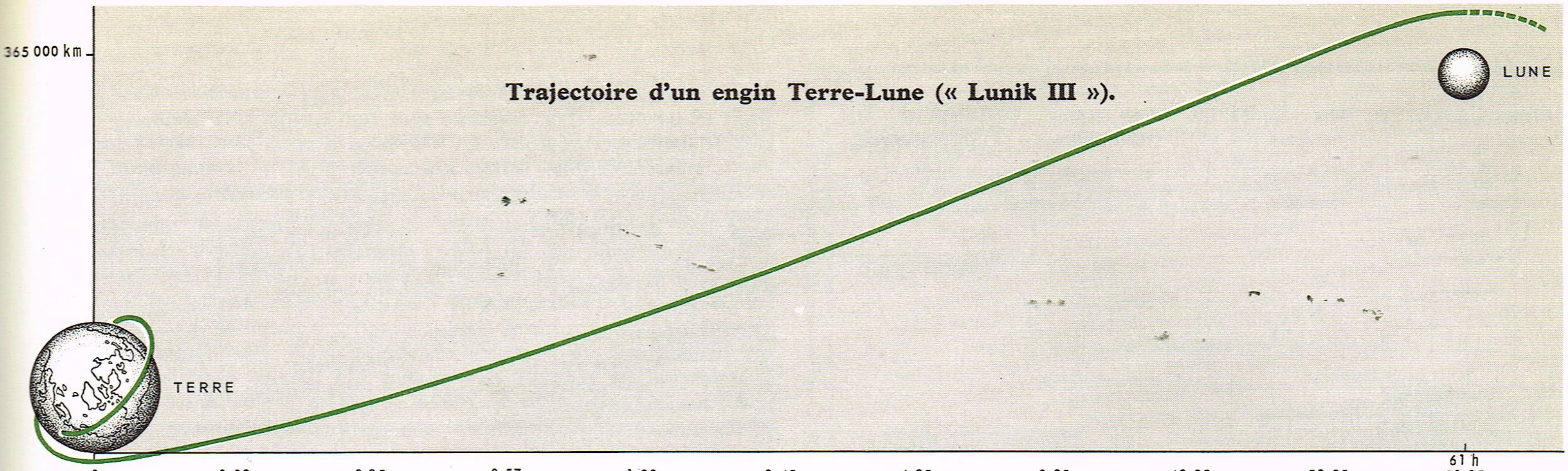
En pratique, ces cas sont très rares, et bien qu'ils fassent la joie de certains bachoteurs et sévissent encore dans certains livres d'exercices, leur intérêt pratique est nul. En fait, une machine électronique à qui on demande de calculer une fonction aussi simple que sin x préfère recalculer une solution particulière de l'équation différentielle $y'' + y = 0$, plutôt que de consulter une table de cette bonne fonction; à plus forte raison préférera-t-elle calculer directement une solution de l'équation différentielle $y'' + y = \mathrm{tg}^2 x$ que de savoir qu'il s'agit de

$$y = A\cos x + B\sin x - 2\sin x \cdot \mathrm{Log}\, tg\left(\frac{x}{2} + \frac{\pi}{4}\right).$$

Cette dernière formule ne nous permet ni de connaître les propriétés de la courbe, beaucoup plus faciles à obtenir en considérant directement l'équation dont elle est la solution, ni d'en avoir la valeur numérique sans passer par le calcul de plusieurs fonctions; c'est-à-dire finalement sans avoir à résoudre plusieurs équations différentielles.

Ici encore, par conséquent, la machine opère de manière totalement différente de celle du calculateur manuel. Il a même été nécessaire d'élaborer plusieurs procédés purement numériques. D'innombrables méthodes ont été proposées, ce qui n'est pas réconfortant : presque toutes sont liées à des classes de problèmes pour lesquelles on les a créées, mais si on a recours à tant de méthodes, c'est parce qu'aucune n'est générale !

Une des premières méthodes consiste, connaissant la valeur de la fonction et de ses premières dérivées au départ, à calculer en un point voisin une valeur approchée (valeur prédite). Le calcul de la dérivée $\frac{dF}{dt}$ à l'aide de la fonction G permet, par une



Trajectoire d'un engin Terre-Lune (« Lunik III »).

durée du parcours depuis le lancement (en heures)	0	0,09	0,25	0,57	1,32	2,45	4,95	9,95	48,95	58,95	61 h 60,95
pas du calcul (en heures)	2^{-7}	2^{-6}	2^{-5}	2^{-4}	2^{-3}	2^{-2}	2^{-1}	1	2^{-1}	2^{-2}	
nombre de points de calcul effectués pour chaque valeur du pas	13	11	10	11	9	10	10	39	20	8	

durée du parcours depuis le lancement (en heures)
pas du calcul (en heures)
nombre de points de calcul effectués pour chaque valeur du pas

nouvelle formule, de corriger cette valeur. Puis on procédera de même pour une valeur plus grande de t. Cette méthode de pas à pas nécessite la définition de l'intervalle de t entre deux points de calcul; le choix de cet intervalle (ou pas du calcul) est fondamental : trop petit, il allonge le nombre d'opérations nécessaires; trop grand, il fait confondre courbe et tangente sur une distance trop grande, ce qui conduit à des erreurs.

Pour éviter ces deux écueils, il y a intérêt à prévoir dès l'analyse d'un problème d'équation différentielle la possibilité de changement automatique du pas, sans intervention de l'opérateur, suivant deux critères bien adaptés : trop faible ou trop forte variation de la fonction et de sa dérivée pour une période s'étendant sur deux ou trois pas; les seuils de déclenchement du changement de pas sont généralement déterminés après essais quantitatifs.

être efficace !

Une fois la « bonne méthode » choisie et éprouvée, un problème d'économie de temps se pose. Où stocker les données de manière que le délai moyen d'attente soit minimal. Ce problème de circulation du flot d'informations est difficile, mais joue un rôle fondamental en gestion et toutes les fois qu'il y a une grande masse de nombres à traiter. De la qualité de la solution dépend non le résultat, mais le coût du calcul.

savoir organiser logiquement

Ce n'est pas tout : l'algorithme choisi, il faut encore organiser les calculs. Cela nécessite l'analyse complète de toutes les possibilités et de tous les incidents susceptibles de se produire : entre la réflexion méthodologique et l'écriture détaillée des opérations, il importe d'établir la logique du programme.

L'ensemble des opérations ne se présente pas de manière séquentielle, mais sous la forme d'un réseau qu'il est commode de représenter dans le plan : suivant le résultat des tests, le flot d'informations se sépare en deux ou plusieurs directions, qui peuvent du reste se recouper ultérieurement. La méthode la plus universelle et la plus sûre pour représenter cette organisation consiste à dresser l'organigramme des fonctions logiques et arithmétiques, c'est-à-dire l'ensemble des chemins à parcourir entre le début et la fin des opérations. En termes mathématiques, on parle de graphe; en termes plus naïfs, de croquis indiquant les différentes étapes du traitement.

Chacun de nous a tracé plus ou moins consciemment des organigrammes, fût-ce de manière inconsciente, lorsqu'il a marqué sur son carnet la liste des rendez-vous et activités de la journée suivante, en tenant compte du fait que le résultat de ses activités matinales influait sur ses activités vespérales : nouveaux rendez-vous, déplacements décommandés...

L'écriture d'un organigramme est purement conventionnelle, et peut être plus ou moins détaillée. Chaque fonction que doit remplir le programme, qu'elle soit élémentaire (lire une information) ou représente un ensemble déjà complexe d'informations (résoudre une équation du second degré ou une équation différentielle), est représentée par un bloc inscrit dans un rectangle; les tests figurent à l'intérieur de rectangles à extrémités arrondies, ou dans des losanges; les regroupements des séquences sont indiqués par des points ou des petits triangles. Si l'organigramme est trop long ou trop enchevêtré, des renvois, numérotés, permettent de découper les différentes phases du traitement : on pourra programmer et éprouver séparément les différentes parties ainsi distinguées, et les ranger dans des zones différentes de la mémoire.

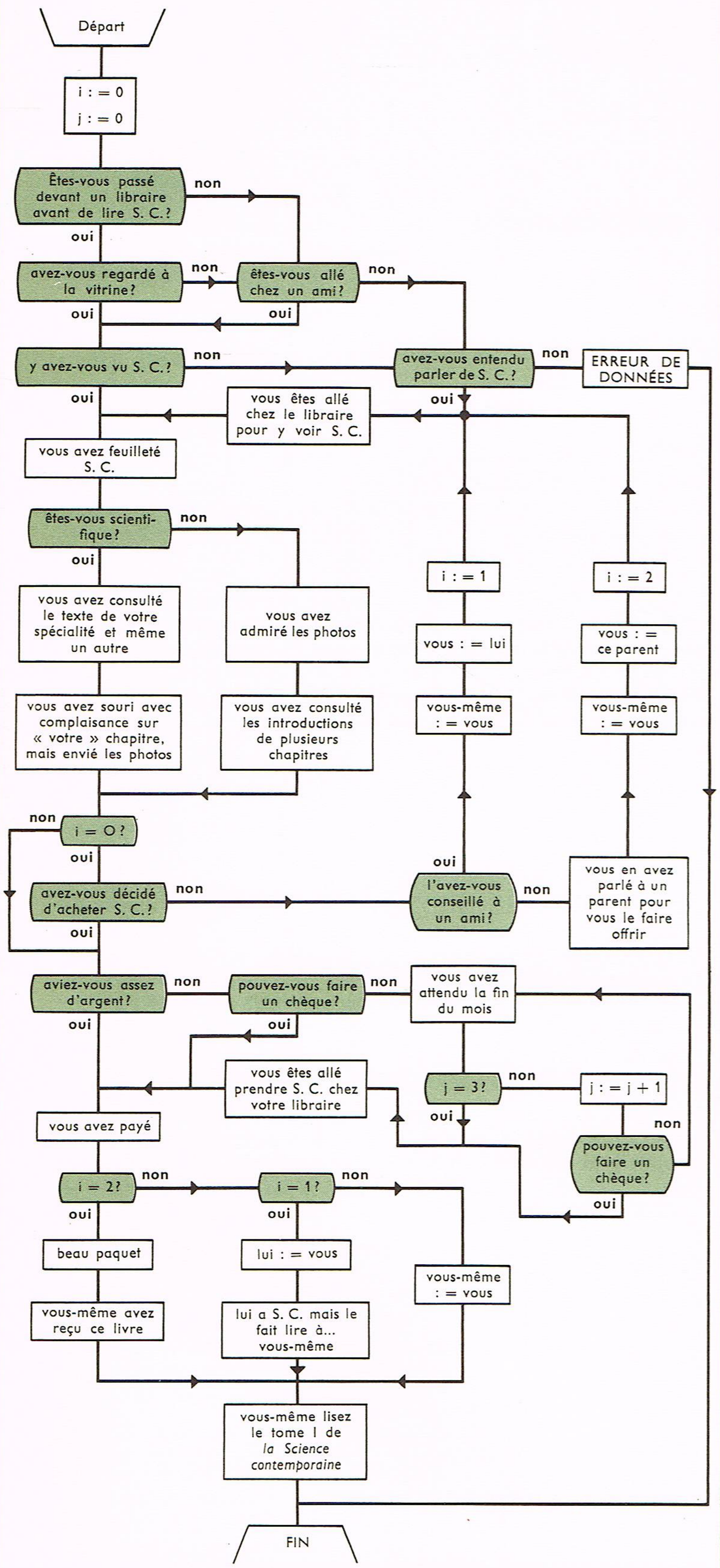
Au cours d'un calcul itératif, un même sous-ensemble d'instructions est utilisé plusieurs fois de suite, certaines valeurs de paramètres étant modifiées d'une étape à l'autre. La portion d'organigramme correspondante s'appelle, improprement d'ailleurs, une boucle.

Il faut évidemment donner à la machine un test lui permettant d'ouvrir un circuit. Si l'on veut adopter un procédé itératif, il faudra indiquer à la machine le moment où il nous semble que décrire une nouvelle boucle n'apporte rien de plus, puisque nous estimerons être suffisamment près du résultat.

Nous avons remarqué dans les organigrammes élémentaires le sigle d'apparence étrange $J := J + 1$.

Cela signifie simplement que la valeur J (figurant dans la mémoire M) doit être remplacée par la valeur $J + 1$: les opérations faisant intervenir le contenu de la mémoire porteront non plus sur la valeur J, mais sur la valeur $J + 1$. Généralement, J est entier; il s'agit alors d'une opération de comptage, et la mémoire M s'appelle *index*.

Organigramme des décisions vous ayant conduit à lire ***la Science contemporaine.***



La règle fondamentale de l'écriture d'un organigramme est finalement la suivante : tout trajet ne peut être parcouru, et toute fonction être exécutée que dans un seul sens. D'autre part, quelle que soit la manière dont une suite d'instructions aiguillent le programme, il doit toujours se terminer par une instruction « fin de programme ».

Une fois l'organigramme tracé, il reste à traduire en langage machine chacune des séquences possibles, en veillant soigneusement à ce que toute séquence amène la machine dans l'état lui permettant d'entreprendre correctement les séquences susceptibles de s'enchaîner à elle, et il faudra notamment enchaîner les instructions de manière convenable. Le déroulement des instructions symbolisées par le sens des flèches exige notamment que toute séquence soit codée, et que les instructions puissent effectivement être prises en charge individuellement, l'une après l'autre, par le registre qui leur est affecté.

L'écriture du programme est aisée, une fois l'organigramme fait et bien fait. Il faut simplement être soigneux, notamment dans la manipulation des données devant intervenir, et dans l'organisation du flot des instructions.

Bien entendu, l'organigramme détaillé tient compte des possibilités effectives du calculateur électronique. L'adaptation d'un programme de 3 000 instructions d'une machine de 8 000 mots à une machine de 32 000 mots est immédiate, et cependant la meilleure utilisation de la machine de grande capacité peut conduire à modifier complètement l'organisation du programme pour profiter des 24 000 mots supplémentaires. On peut, par exemple, introduire en mémoire adressable ce qui était primitivement contenu dans une bande magnétique lue en 48 blocs de 500 mots. L'avantage d'utiliser la nouvelle machine sera double, car on aura réduit à la fois le temps de calcul (vitesse propre supérieure) et les temps d'attente de lecture.

Certaines fonctions sont d'un usage classique; il sera donc économique, quant au temps de programmation, de pouvoir utiliser des séquences stéréotypées pouvant intervenir sur simple appel. On les appelle les sous-programmes généraux, par opposition aux sous-programmes particuliers. Citons, par exemple : les fonctions circulaires sinus, cosinus, tangente et leurs inverses arcsinus, etc.; l'exponentielle e^x et les logarithmes; les fonctions hyperboliques *chx*, *shx*, dites fonctions à une entrée, car elles ne dépendent que d'une variable; ce peuvent être aussi des fonctions de Bessel dépendant de deux entrées, l'argument x et l'entier n, ou des calculs de corrélation, classiques en statistique, portant sur n variables aléatoires.

Les sous-programmes de fonctions indiqués à titre d'exemple sont des séquences utilisées pour calculer la valeur de ces fonctions aussi souvent qu'elles se présentent. Ils remplacent alors les classiques tables de sinus, de logarithmes ou même de fonctions moins connues. Tandis que le calculateur disposant d'une machine de bureau lit dans une table et interpole pour déterminer la valeur utile, le calculateur électronique évite de remplir des dizaines de mètres de bande pour y inscrire les valeurs des sinus de millième en millième de degré (par exemple) entre 0 et 45° : le calcul du sinus exigera seulement l'écriture de quelques instructions en code machine et, compte tenu d'une dizaine d'itérations éventuelles, demandera moins de temps que la recherche dans une table.

une machine ne vaut que par sa bibliothèque de programmes

Certains sous-programmes sont utiles pour une classe d'utilisateurs particuliers, par exemple les centres travaillant sur des problèmes de résistance des matériaux, d'autres pour un très large éventail de disciplines, comme la recherche opérationnelle ou la gestion des entreprises. De la sorte, de même que les bureaux d'études ont des méthodes particulières pour établir leurs projets, les centres de calcul groupent dans une biblio-

thèque les sous-programmes d'usage courant. Cette bibliothèque est constituée par :
— la liste imprimée des fonctions de chaque sous-programme, et même des programmes intervenant fréquemment;
— les explications détaillées sur l'organigramme, et comportant la liste des instructions et aussi le mode d'utilisation pratique (indications du nombre et du type de représentation des quantités nécessaires au fonctionnement du programme et du mode de sortie vers les mémoires de rangement ou vers une imprimante);
— les cartes ou le ruban comprenant toutes les instructions;
— les bandes magnétiques groupant certains ensembles de sous-programmes suivant les problèmes susceptibles de les utiliser.

Comme les sous-programmes deviennent rapidement une sorte de patrimoine nécessaire à tous les utilisateurs d'un même type de calculateur, ceux-ci se réunissent généralement en groupements pour mettre en commun les sous-programmes généraux, de façon à grossir la « programmathèque » valable sur une machine.

De plus, comme cette classe de sous-programmes est appelée à être utilisée très souvent sur beaucoup d'équipements, il y a un intérêt économique évident à rendre ces moyens de travail absolument sûrs, de déroulement aussi rapide que possible et d'emploi le plus simple et le plus souple possible.

langage machine et langage symbolique

Le code des machines est le plus souvent à une seule adresse. Son emploi nécessite le recours à l'écriture de symboles purement numériques sur la feuille devant porter les instructions du programme. Un premier allégement de la tâche du programmeur est obtenu en substituant aux symboles numériques des symboles mnémoniques (ADD pour l'addition au lieu de 10, SOU pour soustraction au lieu de 11). Cette simplification purement formelle est obtenue au moyen d'un programme enregistré auparavant dans la mémoire du calculateur et dont le rôle est de traduire, dès l'entrée des signaux, la partie type d'opération du code alphabétique en code machine suivant une correspondance biunivoque. Le second allégement consiste à remplacer les adresses effectives des données intervenant dans le programme par des adresses relatives dont il est possible de fixer l'origine seulement au moment du passage en machine. De même, pour pouvoir translater à l'intérieur de la mémoire toutes les instructions d'un programme ou d'un sous-programme, et éviter la double utilisation de certains mots, il est indispensable de pouvoir décaler les mémoires où doivent être rangées les instructions dont les adresses interviennent sous forme de nombres dans le compteur d'ordres capable de sauter d'une instruction à une autre. L'ensemble de ces fonctions est rempli par des programmes logiques, appelés généralement « assembleurs » ou « systèmes ».

Cependant, une étape importante a été franchie le jour où l'on a pensé à utiliser des formules globales susceptibles d'être introduites directement dans le calculateur.

La substitution de
Calculer Z : $Ax^3 + 2 (\cos [BT - x]) xy^2$
à la succession des 10 opérations

multiplication	BT
soustraction	$BT - x$
sous-programme cosinus	$\cos [BT - x]$
addition de deux nombres égaux	$2 \cos [BT - x]$
multiplication	$2 (\cos [BT - x]) y$
multiplication	$2 (\cos [BT - x]) y^2$
multiplication	$A x$
multiplication	$A x^2$
addition	$Ax^2 + 2 (\cos [BT - x]) y^2$
multiplication	Z

permet de remplir en une seule ligne une case complète d'un

SOMME MATRICIELLE

1. FONCTION

Il s'agit d'additionner chacun à chacun les M N éléments de 2 tableaux A et B de M lignes et N colonnes pour former un nouveau tableau C de M lignes et N colonnes.

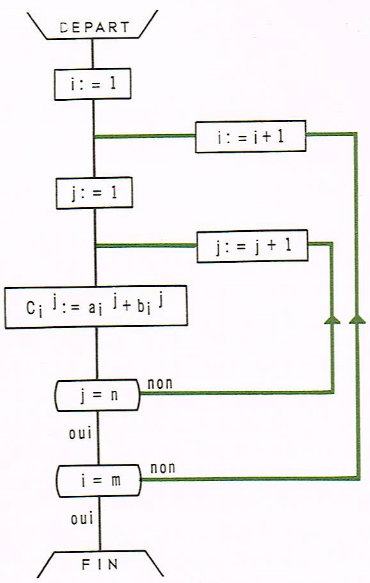
2. EXEMPLE

$$\begin{pmatrix} 1 & 4 & 6 \\ -7 & 0 & 8 \end{pmatrix} + \begin{pmatrix} -2 & 8 & -1 \\ 9 & 4 & -2 \end{pmatrix} = \begin{pmatrix} -1 & 12 & 5 \\ 2 & 4 & 6 \end{pmatrix}$$

M = 2, N = 3, les éléments sont appelés a_i^j, b_i^j, c_i^j

$$C_2^3 = a_2^3 + b_2^3 = 8 - 2 = 6$$

3. ORGANIGRAMME



4. PROGRAMME ECRIT EN LANGAGE MACHINE (CAB-500)

```
T12M113, 1S
F
A8EV()
RV6Δ
T12M113, 1
A8EV1
V4TR6
RV1499ΔY13
V5QR7
RV1499ΔY14
M4R5
T13R4
RCV33ΔI
A7EV()
T1R15
A5EM()+
T2R15
A6EM()+
RV21ΔY8
AF5R6
RV22Δ
SF5R6
T3R15
T5M()+
A7V1
V7TR4
RV29ΔY13
RCV63ΔI
RM113,1
A1V1
A2V1
A3V1
RV14Δ
```

5. PROGRAMME ALGOL

ALGORITHME

procédure SOMME MATRICIELLE (A, B) ***Résultat :*** **(C)** ***type :*** **(M, N);**
réel tableau **A, B, C;** ***entier*** **M, N;**

commentaire **cette procédure calcule les termes C (I, J) d'une matrice C, somme de deux matrices A et B de M lignes et N colonnes, I est l'indice des lignes, J l'indice des colonnes;**

début *entier* **I, J;**

pour **I : = 1** *pas* **1** *jusqu'à* **M** *faire*

pour **J : = 1** *pas* **1** *jusqu'à* **N** *faire*

C (I, J) : = A (I, J) + B (I, J)

fin

organigramme; mais écrire de telles formules exige des règles de construction grammaticale très précises. On appelle *langage* un ensemble de règles de construction de phrases (c'est-à-dire de formules) et les symboles ou mots individuels pouvant être utilisés. Ces mots formant la syntaxe du langage de programmation sont constitués à l'aide des lettres de l'alphabet, des chiffres, des signes opératoires ou logiques et des parenthèses, dont le rôle est d'autant plus important que les formules doivent pouvoir être écrites sur une seule ligne.

Les techniques de fabrication des compilateurs, programmes capables de traduire en code interne, propre à une machine particulière, ont progressé à pas de géant en quelques années. Mais elles se sont heurtées parfois à des difficultés provoquées par le manque de rigueur dans la définition du langage utilisé. C'est pourquoi les langages les plus évolués ont dû être définis par des équipes comportant des mathématiciens spécialistes de logique ou de programmation faisant de la technique grammaticale pour définir les liaisons sémantiques nécessaires à la classe de programmes étudiée. Citons :

— l'Algol (algorithmic oriented language), défini de 1958 à 1960 par un groupe international de mathématiciens en vue d'universaliser un langage scientifique général et sans ambiguïté. L'écriture des formules en Algol permet, dans de nombreux cas pratiques, d'écrire ou de lire un programme sans recourir à l'organigramme. On peut légitimement penser que l'Algol, inclus actuellement au programme des certificats d'études supérieures et enseigné dans nombre d'écoles hautement spécialisées, sera diffusé prochainement dans toutes les écoles d'ingénieurs et, d'ici quelques années, dans les classes terminales des lycées pour initier les élèves aux méthodes de pensée nées des progrès du calcul automatique;

— Le Cobol (common business oriented language), destiné à la manipulation d'ordres de traitement (sélections, interclassement, fichier, lecture ou écriture de libellé) et utilisé par les centres de calcul attachés aux administrations ou aux affaires;

— Le Comit (pour informations non numériques) et le Lisp (pour structures très spécialisées d'instructions dites traitement en listes), créés à propos de questions de traduction et de démonstration automatiques, ont trouvé des débouchés plus larges que ceux initialement prévus.

la mise au point des calculs, dernière étape avant l'exploitation

Après l'écriture du programme, les instructions et un ensemble de données cohérentes sont perforées sur cartes ou sur ruban. On procède alors à un essai de fonctionnement du programme. Parallèlement, des calculs types, effectués par exemple à l'aide d'un calculateur de bureau, permettent de vérifier si le déroulement du programme s'est effectué de façon correcte. Il est généralement impossible de faire emprunter tous les chemins possibles entre le départ et la fin du programme. Par exemple, il y a 136 chemins définis sur l'organigramme de la décision de lire *la Science contemporaine* pour 15 tests seulement; un programme moyen de 10 000 instructions, dont 350 tests, pourrait offrir 6 000 chemins différents. Ce qui est essentiel, c'est de vérifier que tous les carrefours empruntés ont suivi des bonnes directions et que les séquences donnent bien les résultats escomptés. Comme les méthodes de calcul automatique sont, par définition même, complexes, le travail de vérification ne peut être fait ni par un programmeur sans notions de mathématiques ni intelligence au moins grossière de la nature du problème traité, ni par un mathématicien sans connaissances du langage utilisé. Il existe ainsi une quantité de travaux où les connaissances nécessaires exigent beaucoup de métier et un personnel adapté : ni le théoricien pur ni le technicien pur ne peuvent venir à bout de la mise au point d'un nouveau programme. Cela est vrai d'une manière particulièrement significative pour les mises en machine de nouveaux algorithmes mathématiques, où les prévisions théoriques peuvent ne pas être couronnées de succès, par suite d'erreurs dues à la substitution aux valeurs exactes de nombres composés d'une quantité limitée de chiffres, et il y a alors une interpénétration entre le travail de recherches mathématiques et celui de codification.

On met au point les calculs et les traitements par lecture de certains résultats intermédiaires. Ce sont d'abord les sorties normales du programme, mais aussi des extractions supplémentaires programmées spécialement pour suivre l'évolution de certains événements, tels que vitesse de convergence ou nombre d'itérations utilisées. Enfin, une analyse pas à pas des instructions appartenant à des séquences déterminées et l'impression de tout ou partie des mémoires rapides du calculateur à une certaine phase du traitement permettent d'achever les corrections nécessaires avant l'utilisation systématique.

L'exploitation des programmes doit être organisée de manière rationnelle, car il n'est pas permis de perdre beaucoup de secondes dont le prix de revient est de l'ordre d'un franc.

Un système de programmes permet de transférer les instructions dans la mémoire et d'assurer le fonctionnement en continu des opérations avec le minimum d'intervention humaine. Comme le temps de travail de l'unité centrale est ce qui revient de loin le plus cher, cette unité ne doit jamais attendre l'arrivée d'informations : c'est donc les parties périphériques du calculateur qui seront souvent en état d'attente et des ordres écrits par une imprimante rapide ou par une machine à écrire disponible sur le pupitre de contrôle indiqueront à l'opérateur les manipulations à effectuer.

La théorie de la programmation est une branche des mathématiques en plein développement. Elle est fondée essentiellement sur les « mathématiques de nos enfants », que nous verrons au second tome et en particulier sur les théories des ensembles et sur l'algèbre de Boole. Elle utilise la théorie de l'information en vue de déterminer le minimum de tests ou de calculs nécessaires à la réalisation des programmes; la théorie des graphes permet de construire les réseaux d'écoulement des données et des instructions et d'en connaître les propriétés. La théorie des langages permet de dégager l'essentiel de l'accessoire parmi toutes les aides à la programmation construites pour chaque calculateur particulier. L'analyse numérique, enfin, permet de choisir un « bon » algorithme et d'en estimer la valeur.

un ordinateur à chaque étage

La mise au point des calculs pose des problèmes qui ont été longtemps la pierre d'achoppement du développement des ordinateurs, car c'était une affaire de spécialiste. En effet, l'utilisateur d'un ordinateur qui a appris la manipulation du langage ALGOL sans être mathématicien ou informaticien, mais a seulement besoin de mathématique ou d'ordinateur, s'est vu couramment consacrer de nombreuses heures à la mise au point définitive : le centre de calcul est éloigné et il faut s'y rendre; bien équipé pour l'introduction d'un programme à expérimenter, ce centre demande cependant quelque temps entre la prise en charge et l'impression des résultats, car le programme étudié doit être inséré dans la file d'attente de travaux divers et il sera peut-être arrêté dès la troisième ligne d'instructions parce qu'un point virgule n'aura pas été perforé à la bonne place! Il faudra donc attendre le « train » (de travaux) suivant pour savoir si la quatrième ligne d'instructions est convenable. Une nouvelle technique s'est fait jour dans quelques centres universitaires (Paris et Massachusetts). Au lieu de faire perforer à l'avance toutes les instructions, l'ingénieur

ou le physicien va utiliser une sorte de machine à écrire et frapper lui-même les instructions de son programme; une liaison par téléphone reliant la « machine à écrire », ou pupitre, au calculateur lui-même assure au fur et à mesure de l'écriture la traduction en langage-machine du programme. L'erreur, si erreur il y a, sera détectée aussitôt et, dans 80 cas sur 100, le palliatif sera frappé dans la minute par l'ingénieur lui-même. Après détection de la correction syntaxique des phrases du langage symbolique et compilation, l'ordinateur pourra exécuter sur entrées des données. Ainsi un seul passage apparent aura permis la mise au point au lieu des nombreux aller et retour d'antan. Les possibilités de partage d'une grande mémoire entre plusieurs utilisateurs disposant chacun d'un pupitre donnent une très grande souplesse à l'utilisateur, qui mobilise en réalité les fonctions de calcul durant un très court laps de temps, la vitesse de calcul étant tellement plus grande que la vitesse de frappe. L'ordinateur peut alors effectuer simultanément — pour l'échelle humaine — plusieurs mises au point et même un calcul plus long qui aura été à peine perturbé. Ce système est d'autant plus intéressant que les pupitres sont nombreux, et chacun dispose alors en réalité d'un grand ordinateur comme s'il était à son étage à la place du petit pupitre et de son fil de téléphone. Des applications de ce principe ont permis aussi les travaux pratiques pour étudiants qui, à quinze ou trente à la fois, peuvent réaliser rapidement leurs manipulations (Toulouse et Paris).

AUJOURD'HUI ET DEMAIN

Les ordinateurs interviennent et bouleversent nos conceptions dans d'innombrables secteurs : de l'administration des entreprises au dépouillement des photographies prises dans des chambres à bulles, des décisions économiques à l'électro-encéphalogramme, de la documentation aux productions artistiques. Il serait vain de vouloir les citer tous. Mais on ne peut comprendre l'importance de cette évolution sans une certaine intelligence d'au moins quelques applications.

La recherche opérationnelle groupe l'ensemble des disciplines scientifiques permettant de définir les meilleures conditions d'emploi d'un matériel donné, ou de tirer le meilleur parti d'une situation donnée de façon à préparer les directives ou les éléments de décision des responsables. Elle procède d'abord par un inventaire comportant l'analyse de toutes les éventualités pouvant se réaliser, puis par l'étude détaillée du phénomène, qu'il soit économique, statistique, administratif, etc. On pourra utiliser pour cette étude la méthode dite de « simulation ».

On entend par là la représentation du système étudié sur un schéma simplifié que l'on appelle le « modèle », susceptible de modification suivant les paramètres devant intervenir. L'analyse met en évidence des comportements statistiques qu'il faut représenter par des lois portant sur des variables aléatoires, et l'intérêt des méthodes de simulation réside dans la connaissance objective des lois tactiques ou économiques qu'elles permettent d'atteindre.

un problème de recherche opérationnelle

Voici un exemple de simulation économique réalisée sur calculateur.

Le problème concret : des hauts fourneaux et des aciéries ont été fondés il y a quarante ans à proximité de la mer, tout auprès d'un port d'accès excellent, sans écluses et avec une marée faible. Le port comporte quelques places pour les navires en attente, un quai où peuvent accoster trois cargos moyens ou deux grands, des grues, des wagons disposés sur une voie parallèle au quai, enfin un terrain d'entrepôt derrière cette voie. Les arrivages et la production ayant augmenté considérablement, la direction du port a voulu étudier les possibilités d'accroissement de la vitesse et de la capacité de déchargement du port. Les temps d'attente de déchargement des bateaux resteront-ils entre des limites raisonnables sous le régime des nouvelles quantités prévues?

Une étude statistique approfondie portant sur une large période donnait l'évolution de paramètres de déchargement : il fallait considérer trente combinaisons (six tonnages de 2 000 à 16 000 tonnes, et cinq sortes de cales). En fait, la plupart de ces possibilités étaient si peu probables qu'il était inutile de considérer vingt combinaisons. Il fallait tenir compte également du temps de déchargement, variable d'un cargo à l'autre, de la longueur des navires et du type de minerai à décharger. L'ensemble des données de la période de référence correspond en principe à 3 000 combinaisons cargo-minerai, sur lesquelles 80 seulement (soit 3 p. 100) correspondent à plus de 90 p. 100 des cas. Ce sont ces 80 combinaisons seules qui sont simulées, ce qui simplifie le modèle et écarte des solutions non significatives parce que trop particulières. En outre, l'arrivée des bateaux pouvait être considérée comme aléatoire, tout au moins pour la période d'été.

Quand un bateau arrive, la décision de le faire accoster ou non est prise suivant le nombre de cargos en cours de déchargement ou en attente, et la longueur de quai disponible.

A chaque instant, c'est-à-dire à chaque période de la simulation, il faut tenir à jour le registre des cargos à quai, calculer la longueur de quai libre, déterminer le début de la phase de déchargement, puis affecter les grues, compte tenu des disponibilités et de leurs caractéristiques.

Le modèle comporte encore d'autres règles relatives à la répartition et à l'emploi de différents matériels annexes (wagons et machines de raclage).

On obtient finalement les documents suivants :

— une fiche par bateau déchargé indiquant : temps d'attente et de service, et travail de chaque élément du port;

— des bordereaux hebdomadaires récapitulatifs;

— un état du déchargement pour chaque période de quinze minutes permettant d'analyser de près les situations difficiles et de suivre la marche de la simulation. En utilisant les règles de fonctionnement actuelles avec un plus grand nombre de cargos venant se décharger, le modèle a permis de voir l'influence des divers paramètres liés aux installations portuaires proprement dites (allongement des quais, achat de nouvelles grues, etc.), puis le modèle a été transformé en vue du calcul des incidences financières de l'accroissement de capacité du port et de la diminution du temps d'attente et de rotation des navires.

un autobus qu'on n'attend pas

Un autre problème de simulation posé à un centre de calcul était celui du *trafic d'une ligne d'autobus.*

Le transporteur désirait « éponger » un trafic donné (ce qui ne signifie pas constant, car une ville évolue rapidement); deux problèmes se posaient : est-il possible de réduire le nombre de véhicules sur une ligne déficitaire, tout en satisfaisant la demande de l'usager?

La création d'un groupe d'immeubles va-t-elle nécessiter la modification de l'organisation des lignes desservant le quartier intéressé?

Une étude préliminaire a permis de déterminer les données :

— un horaire;

— les lois statistiques d'arrivée des voyageurs aux stations;

— les lois des trajets effectués à partir de chaque station.

Le modèle est correct s'il donne pour la période étudiée le nombre de places occupées à chaque instant dans le véhicule en circulation. Une telle simulation a permis de fournir aux responsables de la compagnie de transport des états d'un premier type donnant le résultat d'une phase particulière de la simulation : pour chaque autobus figure, en face de la liste des stations, le nombre de voyageurs de l'autobus pour les différents circuits de la période de simulation. Des états d'un second type donnent pour les mêmes listes de stations les valeurs moyennes, les écarts types et les coefficients de variation des échantillons de *n* valeurs constituées au cours de *n* phases de traitement.

Comment utiliser ces résultats?

L'horaire en vigueur permet de tester la fidélité du modèle. Pour résoudre le premier des deux problèmes posés, celui de la ligne déficitaire, on établit un nouvel horaire tenant compte de la diminution du nombre de véhicules et on le soumet au modèle en utilisant les paramètres tirés des relevés des enquêtes préliminaires. On remanie cet horaire et on l'essaie de nouveau si nécessaire. Mais il faut procéder à une enquête longue, difficile et coûteuse pour connaître les désirs de la population et dépenser un nombre considérable d'heures-machines pour réaliser la simulation. Le problème est difficile à programmer : on doit notamment tenir compte de ce que la population se sera accrue pendant les deux années qu'il aura fallu pour résoudre le problème initial et par conséquent avoir une analyse très souple de la situation. Malgré ces difficultés, l'avantage de la méthode est ici tout à fait évident. Les essais sur calculateurs ne risquent pas de mécontenter l'usager, comme le feraient des modifications répétées d'horaires perturbant le trafic bien avant de l'améliorer.

De plus, le calculateur peut se permettre d'essayer un très grand nombre de possibilités et donner de surcroît des renseignements auxiliaires intéressants (sur l'opportunité d'utiliser des véhicules de capacité différente, par exemple).

Un problème connexe est celui de la régulation des feux dans une grande ville : Toronto en a bénéficié la première.

du pipe-line à la voie ferrée

Les interventions de la recherche opérationnelle et du calcul automatique sont très fréquentes pour la distribution des produits pétroliers ou le planning des grands chantiers, et le calculateur est utilisé avec fruit pour la détermination de ponts enjambant les autoroutes ou pour le génie chimique. Pour cette dernière branche, citons les calculs de constantes de volatilité, de densités de mélanges d'hydrocarbures, d'échangeurs de chaleur, de condenseurs, de distillation complexe, de recherches sur les réacteurs catalytiques, couramment programmés à l'heure actuelle.



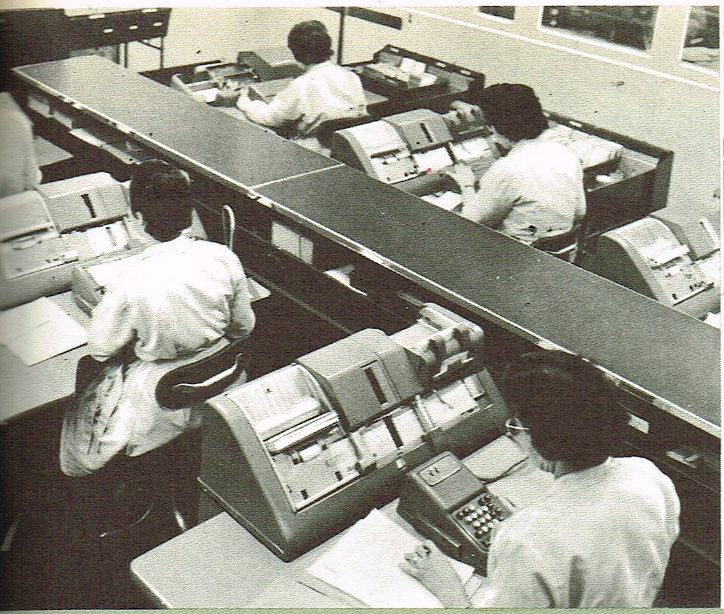
trois ateliers de gestion : machines comptables sans cartes

La *gestion des entreprises* faisait appel il y a quelques années à des ensembles mécanographiques disséminés dans toute la France. Les grands calculateurs ont permis le regroupement de nombreux ateliers en un centre unique, où il est permis de pratiquer le « traitement intégré des données ». Nous citerons le cas de la S. N. C. F., où les données entrent quotidiennement par quantités considérables. C'est ainsi qu'apparaît le bilan 1959 (à l'époque des ateliers disséminés) : 32 milliards de voyageurs-kilomètres et 53 milliards de tonnes-kilomètres de marchandises, 390 millions de trains-kilomètres, 390 000 wagons, 6 000 gares, 370 dépôts et ateliers; le personnel était constitué de 360 000 agents et il fallait régler également 400 000 retraités représentant 1 million et demi de prestataires de la caisse de sécurité sociale gérée par la S. N. C. F. En outre, l'expédition des bagages nécessitait 7 millions de documents d'expédition. Les ateliers, qui employaient 1 200 mécanographes, étaient équipés de 200 machines électromécaniques et d'un petit calculateur. Le traitement au centre unique de la porte d'Auteuil, à Paris, équipé d'un Gamma 60, permet en une minute de lire 600 cartes de 80 caractères, d'en perforer autant, d'imprimer 1 800 lignes de 120 caractères; l'équipement du calculateur comporte 18 dérouleurs de bandes magnétiques et des tambours.

Maintenant, les chefs d'arrondissement de la S. N. C. F. disposent d'un instrument de traitement à leur niveau et la perforation se fait à l'échelon local le plus systématiquement possible, pour éviter de multiples opérations manuelles. Tous les fichiers de gestion sont tenus à jour, les documents de sorties (paies, mandats, ...) sont établis directement. Enfin la Direction générale est informée au jour le jour, grâce à des statistiques, de la situation générale de l'entreprise.

du résumé concis à la traduction toute faite

Le problème de la *documentation* devient absolument capital dans les grands centres de recherche, où l'on reçoit chaque année 200 000 articles ou livres. La consultation d'un tel fichier devient gigantesque et la demande quotidienne de 200 documents ne peut être satisfaite convenablement par les moyens artisanaux dont peut disposer le meilleur personnel documentaliste. C'est pourquoi un gros effort a été tenté ces dernières





mécanographie à cartes classiques — **électroniques (type IBM 1401)**

années en vue de résoudre ce problème. Des différentes investigations tentées, seules peuvent être retenues celles qui font intervenir des systèmes de traitement d'informations susceptibles d'opérer des remises en ordre quotidiennes portant sur 1 000 informations de 50 à 100 caractères à insérer dans la bibliothèque de cartes ou de bandes, tout en permettant la sélection de 200 articles repérables par des indices appelés « mots clefs ». Un problème de nature différente, mais posé pour des considérations analogues, est celui de la création de *résumés de documents* à cadence suffisante. De nombreuses analyses statistiques ont permis de procéder à des essais de méthodes pour faire composer automatiquement des résumés compréhensibles. Des difficultés surgissent lorsque les auteurs utilisent des synonymes pour éviter les répétitions, ce qui peut conduire à définir des tables d'équivalence de volume important. Cependant, quelques expériences ont été couronnées de succès. Les recherches sur la *traduction automatique* ont commencé vers 1957. Dans l'état actuel de la technique, seuls les textes traitant d'un sujet scientifique peuvent être traduits. Cette limitation est due à plusieurs raisons :
— la qualité fondamentale de la traduction scientifique est la fidélité à l'original et non l'élégance de style;
— la rapidité escomptée de la traduction à l'aide des machines électroniques peut apporter une aide fondamentale à la documentation des grands centres de recherche;
— l'intérêt de la défense nationale conduit à subventionner, dans tous les grands pays, les recherches sur la traduction des textes scientifiques.

Un des premiers problèmes est celui du dictionnaire, en liaison avec la syntaxe. Si l'idée de départ était, il y a une demi-douzaine d'années, d'utiliser des lexiques pour effectuer la traduction mot à mot, les premiers essais ont été si peu concluants que les programmes de traduction actuels comportent des recherches syntaxiques en vue de réduire les mots du dictionnaire aux radicaux ou aux formes les plus usuelles. Méditons ces deux exemples pourtant bien simples : les poules du couvent couvent; dans la dure lutte de la course contre la montre, Anquetil ferme la marche. Où le calculateur électronique va-t-il chercher le verbe? En fait, l'œuvre ne peut être réalisée sans le recours aux techniques d'approfondissement lexical syntaxique et sémantique, et une meilleure liaison entre linguistes et mathématiciens en a été le premier fruit. Des modèles de structure de langages ont été tentés et la phase de la traduction directe du russe en français — par exemple — peut être remplacée par l'utilisation de langues intermédiaires. L'emploi d'une métalangue unique est très séduisant, puisque les dictionnaires français-



Première utilisation d'un « Mark I » comme machine à traduire.



« *Spirit is ready, but flesh weak* » (« L'esprit est prompt, mais la chair est faible »). La machine lisant *spirit* et *flesh* pense qu'il s'agit de termes gastronomiques et aboutit à cette curieuse traduction de ce verset : « On a servi un alcool et une viande sans goût. »

Phot. Marmand, I. B. M. Phot. Larousse. (« Le Gourmand », par Boilly.)

métalangue (et vice versa) et russe-métalangue (et vice versa) permettent d'assurer les traductions de et vers le français et le russe. Cependant, une telle métalangue est si lourde qu'il se révèle beaucoup plus fructueux de définir une langue voisine du français et une langue voisine du russe comme langages intermédiaires. Ces langues pouvant servir d'entrée ou de sortie à une langue naturelle sont le lieu d'embranchement normal; de structures parallèles les unes aux autres et utilisant le même vocabulaire que les langues naturelles auxquelles elles sont rattachées, elles permettent donc la traduction en trois phases. Le problème de consultation de dictionnaire est si complexe que certains constructeurs de machines ont poussé leurs investigations jusqu'à la réalisation d'un dictionnaire matériel (et non plus seulement programmé à l'imitation d'une table numérique) par utilisation d'un disque mince (pellicule plastique) dont les informations binaires (noir et blanc) extrêmement denses sont lues à toute vitesse à l'aide d'un bras mobile portant une tête de lecture très fine. Actuellement, ce type de mémoires dites photostatiques préfigure les calculateurs de demain et même d'après-demain par la condensation de l'information, et l'on peut dire que traduction et calculateurs se sont fait réciproquement progresser.

une machine qui sait lire !

Cependant, le vrai goulet d'étranglement de la traduction automatique est l'obligation de transformer les livres ou articles en impressionnants volumes de ruban perforé. La tâche de manœuvre du traitement de l'information appelle aussi une automatisation. C'est pourquoi des recherches sont entreprises (notamment en France à l'Institut Blaise-Pascal du C. N. R. S.), en vue de la réalisation d'une *machine à lire*. Le traitement des informations bancaires a conduit à la création de cartes où les perforations étaient remplacées par des traits au graphite, détectables par des « magnéto-lecteurs », puis plus récemment des caractères magnétiques de forme particulière (selon le code C M C 7) ont pu être lus par des trieuses « liseuses ». Mais l'intérêt est de partir de documents imprimés par des procédés classiques. La technique fait alors appel à une cellule volante se déplaçant verticalement et portée par un support mobile horizontalement en vue de balayer les caractères d'une ligne pour simuler l'opération de lecture faite par l'œil. Un balayage toutes les millisecondes permettra d'atteindre des performances de lecture supérieures à celles de l'œil, et donc considérablement supérieures à la perforation-vérification de cartes, qui nécessite au moins deux lectures entre les opérations tactiles. Il est plus difficile de lire les caractères italiques que les caractères latins, car l'*y*, par exemple, présente des noirs au-dessous du caractère précédent et le *t* au-dessus du caractère suivant : la définition des noirs et des blancs d'une lettre ne peut être achevée avant la suivante et des considérations de connexité (continuité et d'un seul tenant) sont nécessaires.

La théorie de l'information permet de définir le nombre minimal de cellules et de tests nécessaires à l'identification d'un caractère : 23 tests, pour un caractère, par exemple (on peut compter 512 caractères différents dans un livre usuel); le déplacement de la cellule sera contrôlé avec un système optique permettant la précision de la seconde d'arc.

Les techniques pouvant permettre d'aboutir à la machine à lire sont fondées sur une liaison étroite entre l'optique, la mécanique et l'électronique, et c'est à un calculateur électronique qu'il appartiendra de reconstituer par programme les caractères du document à lire. Des méthodes particulières de programmation doivent être développées pour permettre la réalisation sur calculateur rapide des travaux d'identification des signes; ces procédés, appelés « reconnaissance des structures », font appel aux capacités logiques des calculateurs, afin de doter ceux-ci de l'équivalent du sens de la vue.

de l'exploration de l'infiniment petit...

Les applications de la reconnaissance des structures semblent extrêmement prometteuses. L'une d'elles vient d'entrer dans la vie courante des laboratoires de physique travaillant auprès des grands accélérateurs. Les centaines de milliers de photographies de chambres à bulles prises chaque année dans des grands centres comme le CERN ne peuvent être exploitées par les méthodes classiques : les appareils d'optique permettant le dépouillement et les instruments de mesure optiques et électromécaniques — desservis par un technicien chargé de faire des mesures sur des agrandissements de photographies de trajectoires nucléaires — étaient conçus pour l'interprétation de quelques milliers de mesures annuelles par laboratoire. L'examen de dix à cent fois plus de photographies par an ne pouvait être envisagé par multiplication pure et simple de ces appareils, déjà saturés. C'est pourquoi un dispositif de cellule volante, spécialement conçu pour l'introduction dans la mémoire d'un calculateur des informations contenues sur pellicule photographique, vient-il d'être mis au point dans les principaux laboratoires travaillant avec des chambres à traces (en particulier : Faculté des Sciences de Paris et Saclay). Une fois l'entrée des données assurée par ce dispositif, des programmes de reconnaissance de structures particulières permettent la sélection des traces les plus intéressantes en vue du traitement en calculateur, par les méthodes déjà plus anciennes, de la reconstitution du phénomène physique enregistré.

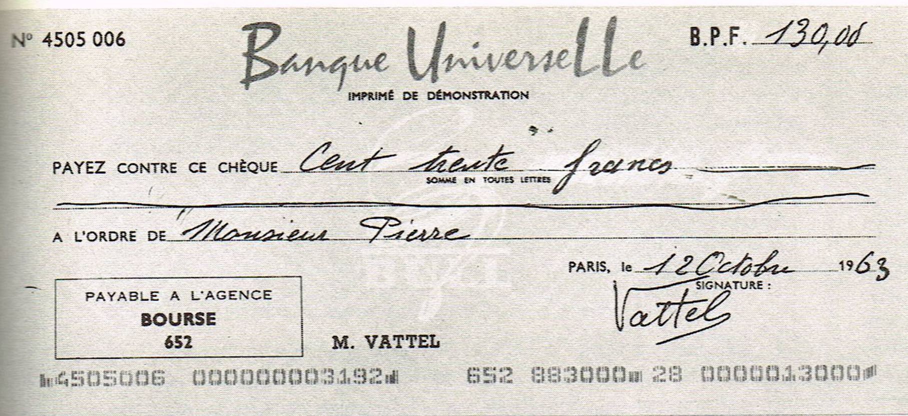
...à celle de l'activité du cerveau

C'est encore par l'analyse des structures que pourra être poussée plus avant la connaissance de notre cerveau.

L'électro-encéphalogramme, c'est-à-dire la mesure électrique de l'activité du cerveau, consiste en courbes donnant le potentiel électrique émis par certaines régions du cerveau. Des électrodes capables de détecter une différence de potentiel d'un microvolt sont placées à l'extérieur de la boîte crânienne; dans certains cas on a réussi à introduire des aiguilles extrêmement fines à l'intérieur du cerveau. En plaçant un nombre important d'électrodes bien réparties (dont on note les cotes) sur la surface du cerveau, on peut reconstituer une carte de l'activité cérébrale, à la manière des courbes de niveau des cartes de géographie. Une autre méthode conduit à placer les électrodes sur une seule ligne et à tracer un schéma coté en microvolts obtenu par développement de la ligne de mesure en fonction du temps. Au cours d'une telle expérience, on peut introduire un facteur extérieur susceptible de modifier l'activité du sujet examiné : sonnerie ou raies de lumière. La corrélation entre le facteur physique et la variation de l'activité cérébrale permet de connaître l'état du patient.

La difficulté de l'analyse des corrélations entre ces facteurs divers et la réaction du cerveau tient au nombre énorme de variables entrant en jeu et à l'importance des travaux de dépouillement. C'est pourquoi les médecins se sont adressés au calculateur électronique, pour arriver à dominer le monceau de résultats bruts des mesures électriques. La réalisation de l'électro-encéphalogramme se fait par couplage direct de l'appareillage de mesure et du calculateur dont le travail est tellement important que les résultats d'une seule mesure sont obtenus en un temps encore excessivement long. L'interprétation par le calculateur des activités cérébrales va jouer un rôle décisif dans certaines interventions chirurgicales, le jour où le temps de traitement aura pu être réduit de façon à obtenir au bout de 1 seconde environ la carte portant sur un phénomène de l'ordre de la milliseconde.

Le facteur de rapidité est encore supérieur à 100 par rapport aux machines les plus récentes. D'ores et déjà l'utilisation de



N° 4505 006

Banque Universelle

IMPRIMÉ DE DÉMONSTRATION

B.P.F. 130,00

PAYEZ CONTRE CE CHÈQUE Cent trente francs

SOMME EN TOUTES LETTRES

A L'ORDRE DE Monsieur Pierre

PARIS, le 12 Octobre 1963

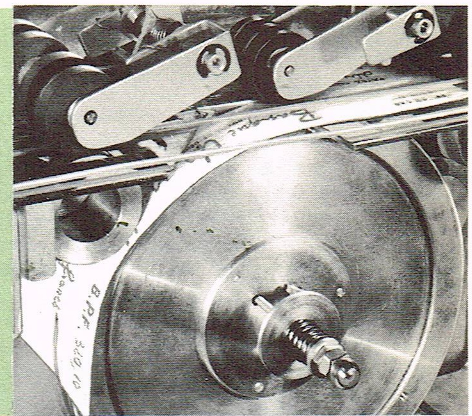
SIGNATURE : Vattel

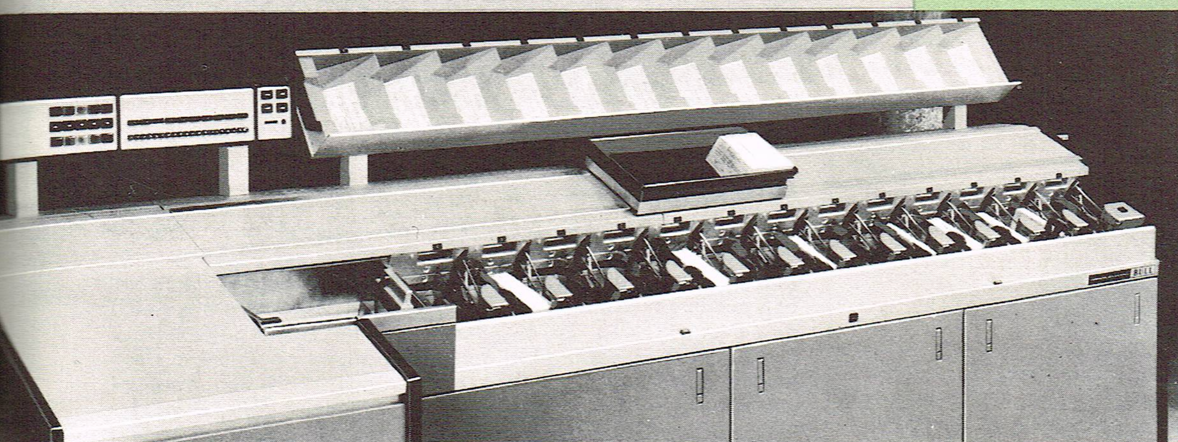
PAYABLE A L'AGENCE
BOURSE
652

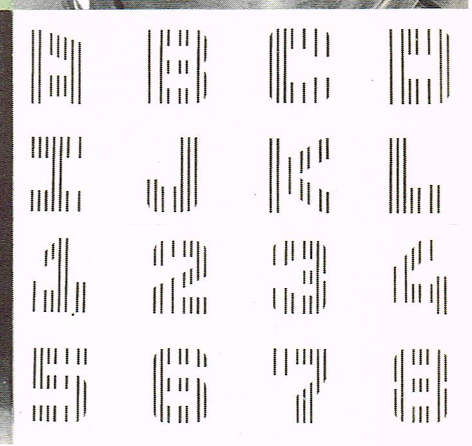
M. VATTEL

4505006 000000003192 652 883000 28 0000013000

La machine CMC 7 sait lire et trier les chèques au bas desquels on a composé des chiffres et des lettres en traits juxtaposés.







l'électro-encéphalogramme a permis un approfondissement sérieux des méthodes par l'exploitation rationnelle de l'information tirée de la mesure électrique de l'activité du cerveau.

l'information, dès qu'elle arrive et d'où qu'elle provienne

Les calculateurs électroniques sont déjà arrivés à un degré de perfectionnement très élevé. Peut-on prévoir aujourd'hui quelles applications nouvelles seront découvertes d'ici seulement cinq ans? Bien imprudent qui peut répondre.

En premier lieu, la sécurité de fonctionnement, l'absence de panne, permettent de se fier toujours davantage aux résultats des calculateurs que n'altère plus l'échauffement local d'une diode. Un des problèmes les plus importants est celui du calcul en temps réel, c'est-à-dire la possibilité de recevoir un ensemble de signaux, de les analyser, de les interpréter et de faire parvenir la réponse pour donner un ordre ou une mesure qui permettra au système émetteur d'agir en fonction de cette réponse. Voici quelques exemples :
— guidage d'une fusée porteuse de satellite pendant les vingt premières secondes de sa trajectoire;
— dépouillement complet d'une photographie de chambre à bulles avant la fin de la série de photographies en cours (trois photos toutes les secondes, et cela pendant plus d'une heure);
— guidage d'une machine-outil.

Le calculateur électronique est alors couplé à un appareillage de transmission d'informations et sert de relais pour la modification ou le contrôle d'un événement sur le point de se réaliser.

Des calculateurs d'un principe totalement différent avaient été utilisés en exclusivité dans ce domaine entre les années 50 et 60 : les calculateurs analogiques. Ces calculateurs spéciaux sont caractérisés par un temps de réponse très bref, car leurs circuits sont calqués sur un phénomène physique analogue au phénomène réel et se passant dans la même durée, mais la précision obtenue ne dépasse pas 3 ou 4 chiffres significatifs. La rapidité des nouvelles générations de calculateurs semble prouver que ramener tous les phénomènes à des successions d'opérations arithmétiques ou logiques, puis reconvertir les résultats en un signal électrique, constitue le système de contrôle le plus rapide et le plus précis, donc le plus efficace.

Pour augmenter la puissance des calculateurs électroniques partout se développe l'utilisation de systèmes de transmission rapide. L'arrivée d'informations peut se faire non plus seulement par cartes ou rubans perforés ou par bande magnétique, mais par téléphone classique et même par l'intermédiaire de satellites du type « Telstar ». Dans ces cas la vitesse d'entrée et donc le débit des signaux sont encore loin d'épuiser la capacité d'admission de l'unité centrale, mais les utilisations de systèmes de traitement à distance sont nombreuses : depuis le contrôle de gestion d'un organisme comme la S. N. C. F., jusqu'à la surveillance de centrales d'énergie électrique, thermique ou nucléaire, à l'aide d'un calculateur situé dans un P. C. éloigné du lieu de production de l'information.

La communication directe entre deux grands calculateurs devient alors un problème sérieux, n'ayant rien à voir avec les échanges entre robots de la science-fiction. Depuis longtemps, certains calculateurs de série sont livrés avec des « satellites » qui sont des machines plus petites, spécialisées dans un travail d'entrée-sortie et de calculs légers, échangeant des informations avec le calculateur central. Pour rendre efficace ce genre de liaisons multiples, les unités centrales doivent pouvoir interrompre instantanément le calcul en cours pour répondre aux appels d'urgence. Ces appels, réglés à l'aide de mémoires spécialisées dans un véritable office de tampon, utilisent l'unité centrale pendant un temps très court (de l'ordre de la milliseconde par exemple), et celle-ci peut reprendre le travail principal avec un retard insignifiant.

Phot. Cie des Machines Bull.

plusieurs choses à la fois

Les difficultés de la programmation de plusieurs problèmes pouvant être traités simultanément et celles du *partage du temps* de l'unité de calcul sont résolues pour les très grandes machines qui commencent à jouer un rôle-pivot dans l'exploitation de quelques grands centres scientifiques (N. A. S. A., CERN, plus généralement énergie nucléaire et expériences aérospatiales), administratifs (chemins de fer, assurances) ou opérationnels. A cette dernière catégorie se rattachent tout spécialement les systèmes de location de places dans les avions. Une compagnie de navigation aérienne qui, sans faire attendre un client placé à Toulouse peut lui dire si le vol Mexico-Buenos Aires du premier jeudi du mois suivant a encore une place aura gagné un passager sérieux. Pour être sûr de la location, le passager a le droit de demander au calculateur une réponse indiquant son nom, le numéro du vol, le type d'avion, le numéro de la place et la classe. Ces systèmes sont utilisés couramment par les agents des compagnies Air France, Panam et Eastern Airlines; agents et clients se rendent-ils compte de combien de progrès techniques la transmission de la demande, la recherche de la location et la transmission de la réponse en une seule communication téléphonique ou radiotéléphonique ou par télex, sont redevables?

des postes clés

Tout le monde sait aussi que le système d'alerte permanent des Etats-Unis fonctionne depuis 1959 à l'aide d'un seul groupe de calculateurs, qui occupe, à vrai dire, un bâtiment de quatre étages.

La sécurité absolue est obtenue par le couplage de deux calculateurs centraux identiques, dont l'un fonctionne pendant que l'autre se fait reviser ou se met en état de veille, prêt à recevoir les ordres de remise en état normal dans les vingt secondes. Les utilisations des calculateurs dans le domaine stratégique ne se bornent pas là, car les méthodes de simulation permettent, depuis de nombreuses années déjà, de déterminer les meilleures utilisations opérationnelles des armes nouvelles, en présence d'un adversaire doté d'armements plus ou moins puissants. Les résultats obtenus ont sans doute déjà permis de réduire les incertitudes de la guerre totale, partant les guerres elles-mêmes.

Les gouvernements disposent de moyens obtenus à partir de connaissances statistiques considérables et voici que certains Parlements demandent l'acquisition de calculateurs qui permettraient de déterminer scientifiquement les conséquences économiques de tel projet et de se prononcer en toute connaissance sur l'opportunité de l'exécution du projet.

Cependant, l'économie des centres de calcul impose une rigidité d'exploitation qui paraît déconcertante aux gens de l'extérieur : le prix de revient des calculs — 2 F la seconde, vingt-quatre heures sur vingt-quatre — impose un rythme auquel on ne peut se plier sans exiger l'automatisation maximale de toutes les entrées ou sorties d'informations. Malheureusement, ce coût des machines (lié à la très grande technicité de tout le personnel d'études et de fabrication) en réserve l'utilisation aux économiquement forts : et les pays en voie de développement sont encore loin de pouvoir profiter des services de traitement de l'information, à part de rares essais de coopération internationale sur l'exploitation en commun des recensements des possibilités agraires ou des industries extractives.

et même l'art

Les sciences et les fonctions administrative, publique et politique ne sont pas les seules à pouvoir bénéficier des services d'un centre de calcul. Théologie et littérature ne sont pas à l'abri de la contagion : le recensement du vocabulaire et de la syntaxe de certains textes permet d'authentifier un document, voire d'en deviner les blancs (manuscrits de Qoumrân).

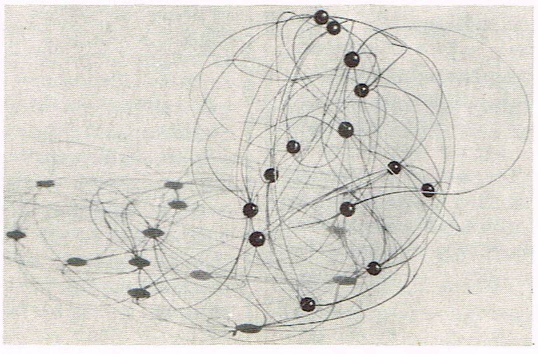
On parle beaucoup de musique et de peinture algorithmiques. De quoi s'agit-il exactement?

Les musiciens utilisent le système de numération de base 12 et repèrent un son par deux chiffres : celui des unités pour la note et celui des douzaines pour l'octave (7 touches blanches et 5 noires sur le piano); par exemple ut_{-1} est noté par convention 0 et sol_0 : 17. En musique, l'addition est pratiquée sous le nom de transposition; la multiplication, en revanche, n'avait encore jamais été utilisée par les compositeurs. La musique algorithmique utilise aussi d'autres opérations dont les modalités choisies aléatoirement dans un certain ordre permettent d'obtenir une polyphonie déterminée par un ensemble de règles. Nous avons d'ailleurs entendu de la musique algorithmique sans le savoir : elle accompagne déjà nombre de films cinématographiques et de grandes œuvres symphoniques ont même été jouées en concert.

La peinture algorithmique fait elle aussi appel aux possibilités offertes par le calculateur électronique. Tout dessin peut être considéré comme un ensemble de points liés entre eux par des lignes, donc comme un « graphe » qui peut être traduit sous forme de matrice carrée, donc « entrer en machine ».

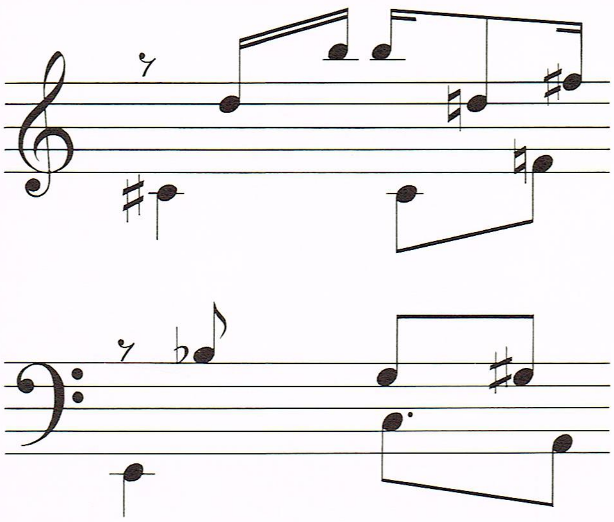
La structure d'une peinture peut aussi être considérée comme un graphe : un centre d'intérêt sera assimilé à un sommet et lié à tous les centres sur lesquels il exerce une influence. Cette méthode permet d'analyser les œuvres existantes, de trouver les lois de composition (au sens de l'esthétique) utilisées par les peintres et de créer de nouvelles œuvres « à la manière de »... Elle permet aussi, ce qui est plus intéressant, de créer une esthétique nouvelle prenant ses sources dans les concepts et les méthodes de la mathématique moderne.

Des expositions de toiles réalisées à partir de schémas et de structures algorithmiques ont déjà eu lieu à Paris et à l'étranger.

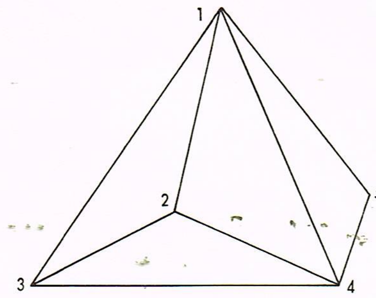


Graphe.

10	10	3	« onze »	·	2	·	6
1	·	·	·	·	·	5	·
10	10	« dix »	·	7	·	8	·
4	·	·	·	0	·	·	9



« Traduction » du tableau fourni par le calculateur (exemple tiré de P. Barbaud et R. Blanchard).

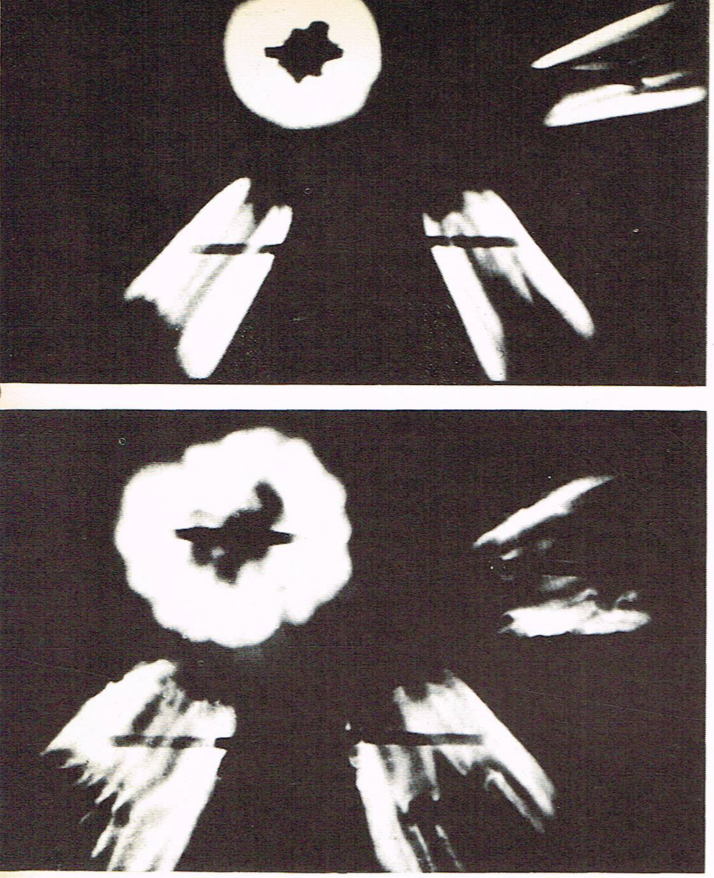


	1	2	3	4	5
1	0	1	1	1	1
2	1	0	1	1	0
3	1	1	0	1	0
4	1	1	1	0	1
5	1	0	0	1	0

Il y a un 1 à l'intersection de la ligne *i* et de la colonne *j* si une ligne (arête) lie *i* et *j* (appelés « sommets du graphe »).

Tous ces domaines de l'activité humaine où a pénétré le calculateur ne peuvent plus évoluer comme auparavant. Le développement des moyens automatiques de l'information exige l'adaptation de milliers de personnes à l'emploi efficace des méthodes modernes. Le métier de mathématicien dans sa tour d'ivoire est tout aussi nécessaire qu'à l'époque de Cauchy ou de Boole il y a cent ou cent cinquante ans; mais le mathématicien appliqué est et sera de plus en plus recherché : à quoi serviraient les nouveaux perfectionnements réalisés par les électroniciens si leur matériel inerte n'était pas conduit par l'homme chargé de rédiger les programmes? Il ne faut pas oublier que les plus perfectionnés des calculateurs sont incapables de rien déterminer tant que l'homme ne leur a pas préparé leur travail, dicté les ordres qu'ils auront à exécuter. Plus les calculateurs se développent et plus il est nécessaire au maître de s'organiser pour obliger la machine à faire un travail qui n'a d'intelligent que la vitesse de déroulement. C'est à l'homme de poser le problème et d'indiquer la méthode de résolution.

Depuis la perforeuse jusqu'à l'opérateur sur calculateur, depuis le programmeur en code machine jusqu'au mathématicien, du technicien d'entretien à l'ingénieur qui crée, du chercheur en laboratoire jusqu'au critique qui interprète les résultats, les carrières liées au développement des calculateurs électroniques offrent un éventail de choix chaque jour plus grand. Des vocations peuvent se donner librement à tout ce qui est attaché aux calculateurs, car le progrès est désormais lié à la façon dont l'homme sait commander à ces aides forgés par lui pour le meilleur et non pour le pire.



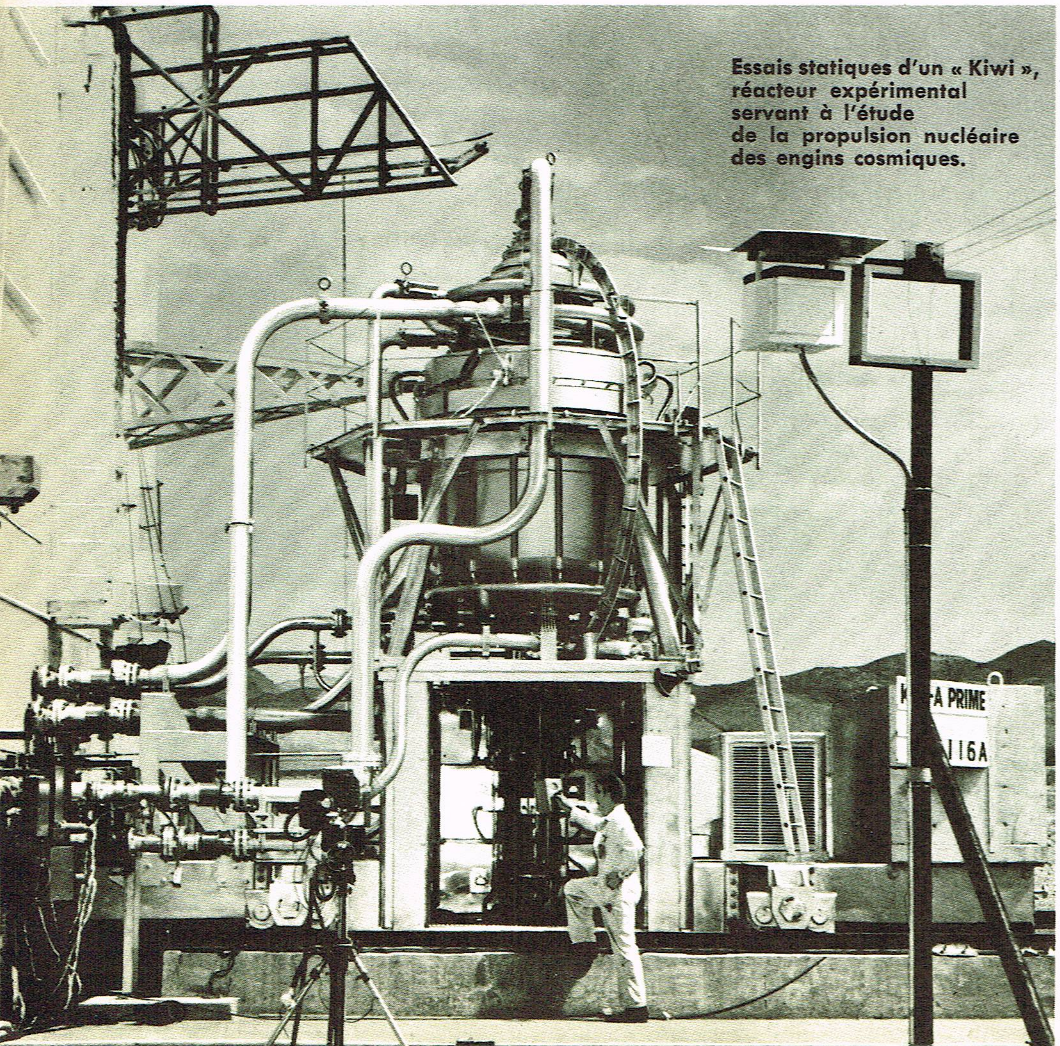
Photographies prises au dix-millionième de seconde lors d'expériences sur la striction d'un plasma. *Ci-contre*, les particules sont confinées au centre du récipient, mais cet état instable ne durera pas et le plasma se répandra de nouveau dans la cavité. (Les trois traînées allongées sont des vues du plasma confiné réfléchies par des miroirs.)

trois jeunes industries

SCIENCE, techniques et développement industriel sont en symbiose étroite. De cette vérité essentielle portent déjà témoignage tous les chapitres précédents, où nous avons vu un progrès technologique ou expérimental aboutir à de nouveaux instruments de mesure, donc à de nouvelles connaissances; et, réciproquement, chaque progrès de la théorie se traduit par l'élaboration d'instruments nouveaux, accroissant le pouvoir de l'homme. C'est un peu par cette interaction que peuvent se résumer l'histoire des grands accélérateurs nucléaires, celle de la radio-astronomie et celle de l'astronautique, sans compter l'essor tout aussi prodigieux de l'électronique.

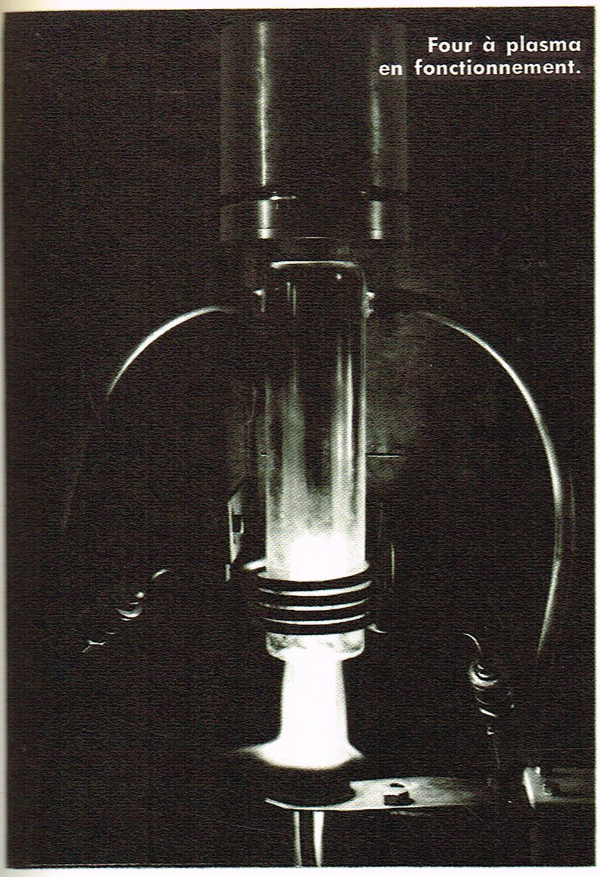
Il nous a cependant semblé bon d'en donner encore quelques exemples, pris parmi les industries de pointe qui réalisent le mieux cette aventure scientifique et technique dans laquelle est plongé le monde, qui influent le plus fortement sur notre vie quotidienne et qui probablement influeront encore plus dans les vingt années à venir sur l'avenir même de l'humanité. Lesquels choisir?

Nous avons déjà montré comment la soif d'énergie qui caractérise une humanité en pleine expansion donna lieu à une véritable « course contre la montre »; sans cesse, il fallait inventer et exploiter de nouvelles sources d'énergie pour remplacer celles que l'homme épuisait de plus en plus vite. Mais voici que sont maintenant mises à contribution ces sources illimitées que représentent l'énergie nucléaire et l'énergie solaire. C'est à la découverte progressive des techniques correspondantes que nous allons assister tout au long des deux premiers chapitres. Nous y verrons tour à tour les problèmes qui se sont posés, et ceux qui se posent encore. Certains, purement scientifiques, font appel à des mathématiques très élaborées ou à des techniques expérimentales très fines : leur étude sera renvoyée au second tome; d'autres, plus techniques, concernent la métallurgie des métaux les plus rares et la maîtrise des températures de plus en plus élevées qui sont en jeu. Non moins essentiels sont les problèmes économiques et politiques que pose l'énergie nucléaire. Il nous faut ici admirer qu'une discipline dont la carrière a commencé à Hiroshima ait fini par ouvrir les plus larges perspectives de collaboration internationale et de paix, en permettant aux pays en voie de développement industriel un accès libéral à une énergie partout disponible à bon marché : les

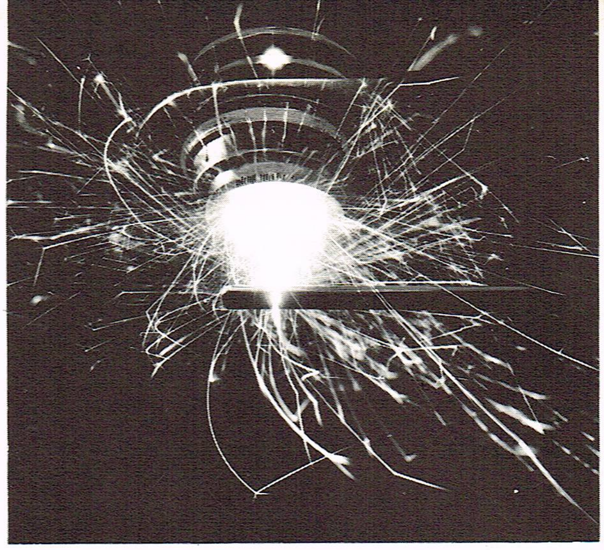


Essais statiques d'un « Kiwi », réacteur expérimental servant à l'étude de la propulsion nucléaire des engins cosmiques.

Phot. Gen. Atomic div., Bru (C.S.F.), U.S.I.S.

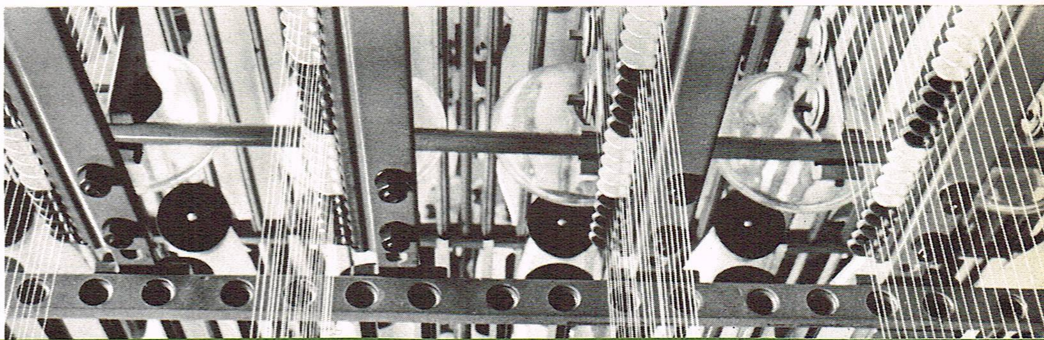


Four à plasma en fonctionnement.

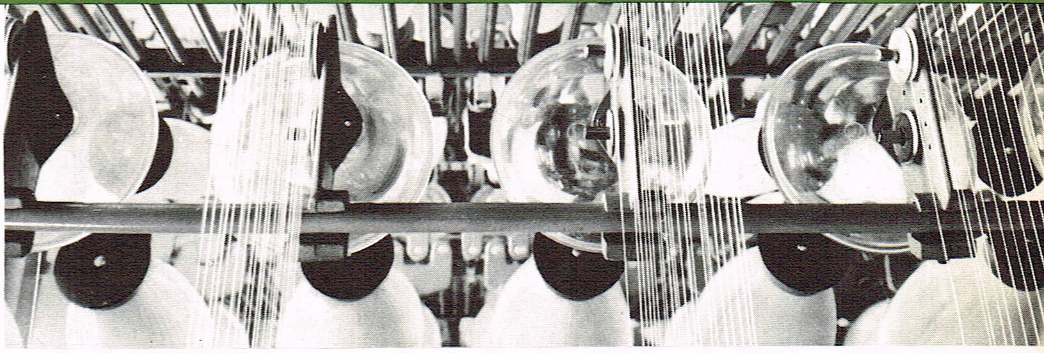


Un éclair de 400 microsecondes produit par un laser à rubis perce une plaque d'acier de 3 mm d'épaisseur. Les traînées lumineuses ont été impressionnées par de minuscules gouttelettes de métal en fusion.

Filage de résines synthétiques.



hautement scientifiques



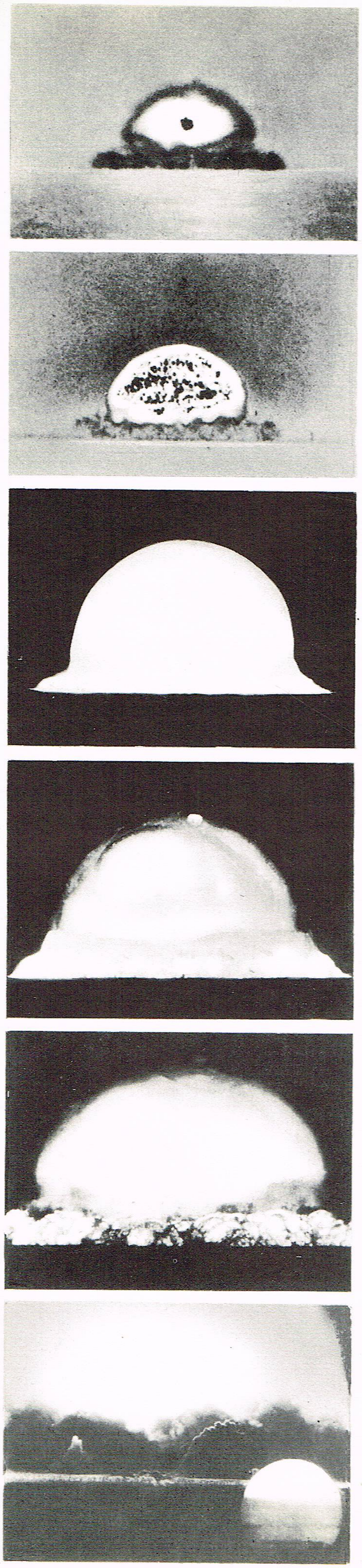
richesses du sous-sol ne deviennent plus le facteur essentiel de la puissance, rôle maintenant dévolu au potentiel humain et aux possibilités de promotion sociale chiffrées par le nombre des chercheurs, des enseignants, des ingénieurs et des cadres, et au dynamisme des fonctions publiques et politiques.

Les hautes températures peuvent paraître un sujet un peu étriqué à côté du précédent. Il n'en est rien. L'exploitation de l'énergie solaire — que ce soit dans les grands fours scientifiques comme celui de Mont-Louis, dans la petite cuisinière que pourra bientôt acheter toute ménagère indienne, ou dans ces usines à rendre l'eau potable grâce auxquelles les déserts redeviendront habitables — se révèle également de grande importance pour un avenir harmonieux de l'humanité. Beaucoup plus que la synthèse du diamant à laquelle nous avons cru amusant de consacrer quelques pages.

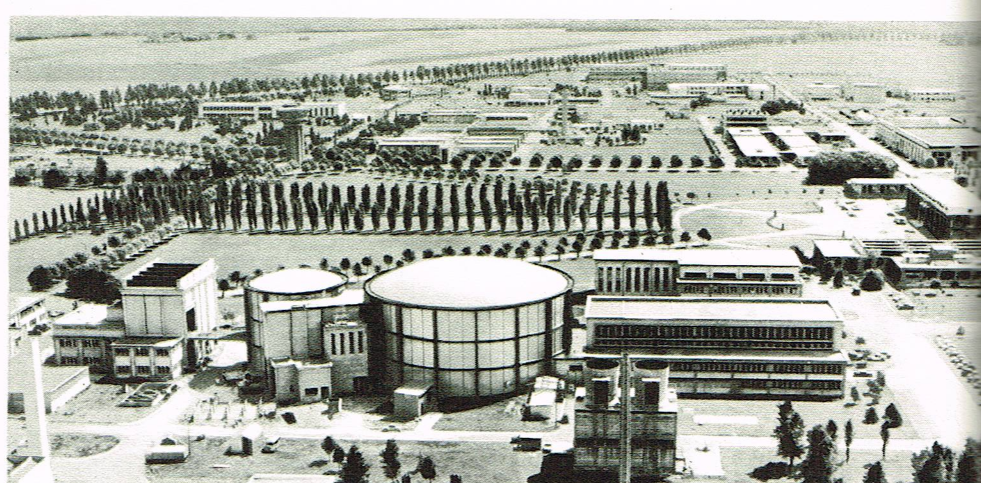
Les matières plastiques jouent un rôle important dans notre vie quotidienne : les industries qui les concernent ont des chiffres d'affaire de plus en plus imposants et consacrent une partie croissante de leurs budgets à la recherche scientifique. Rien d'étonnant à cela; nous sommes ici très près des théories les plus fondamentales de la science moderne, puisqu'il s'agit essentiellement de connaître les modalités selon lesquelles des particules élémentaires comme les atomes s'unissent — et unissent leurs électrons périphériques — pour former les objets qui nous entourent. Le chapitre sur les macromolécules sera donc une excellente introduction à celui que nous consacrons dans le tome II à la « mécanique quantique », et surtout à son complément direct, l'« architecture de la matière ». De plus, les macromolécules jouant un rôle essentiel dans les processus vitaux, il nous a semblé impossible de ne pas parler de l'information génétique et de son support, l'« A. D. N. ». Nous avons dû nous borner ici à mentionner quelques-uns des problèmes scientifiques et techniques ainsi posés, renvoyant le lecteur plus curieux à l'ouvrage *la Vie,* qui est paru dans la même collection.

Phot. Laboratoire central de télécommunications, Rhodiaceta.

ÉNERGIE ET



Ce film marque le début fracassant de l'ère atomique; il a été pris au Nouveau-Mexique en juillet 1945, lors de l'explosion de la première bombe A.



Depuis lors, des perspectives plus intéressantes ont été ouvertes à l'énergie atomique par la recherche nucléaire, dont nous voyons ici l'un des hauts lieux, le centre de Saclay.

Phot. Jean Biaugeaud (C. E. A.).

INDUSTRIE NUCLÉAIRES

par
Georges PARREINS

Le 6 août 1945, depuis le navire qui le ramenait aux Etats-Unis, après la conférence de Potsdam, le président Truman révélait lui-même au monde l'existence d'une nouvelle bombe qui avait été lancée quelques heures auparavant sur la ville japonaise d'Hiroshima.

Les premiers communiqués de presse indiquaient des dégâts immenses, plusieurs centaines de milliers de victimes, plusieurs kilomètres carrés de la ville entièrement ravagés par l'incendie ou détruits par le souffle de la bombe. Le monde entier fut stupéfié et personne pourtant ne se doutait encore de l'importance exacte de ce sinistre. Puis une nouvelle terminologie fut abondamment utilisée pour expliquer comment fonctionnait cette nouvelle arme : neutron, fission, fusion, plutonium, masse critique, bombe atomique, énergie nucléaire... Que signifient exactement tous ces nouveaux vocables?

Qu'est-ce que cette nouvelle énergie que l'on qualifie tantôt d'atomique, tantôt de nucléaire? En quoi diffère-t-elle des autres, comment peut-on la libérer? Peut-on l'utiliser d'une manière moins fracassante? Peut-elle rendre service aux hommes? Ce sont autant de questions auxquelles nous allons essayer de répondre dans les pages suivantes.

le noyau, source d'énergie

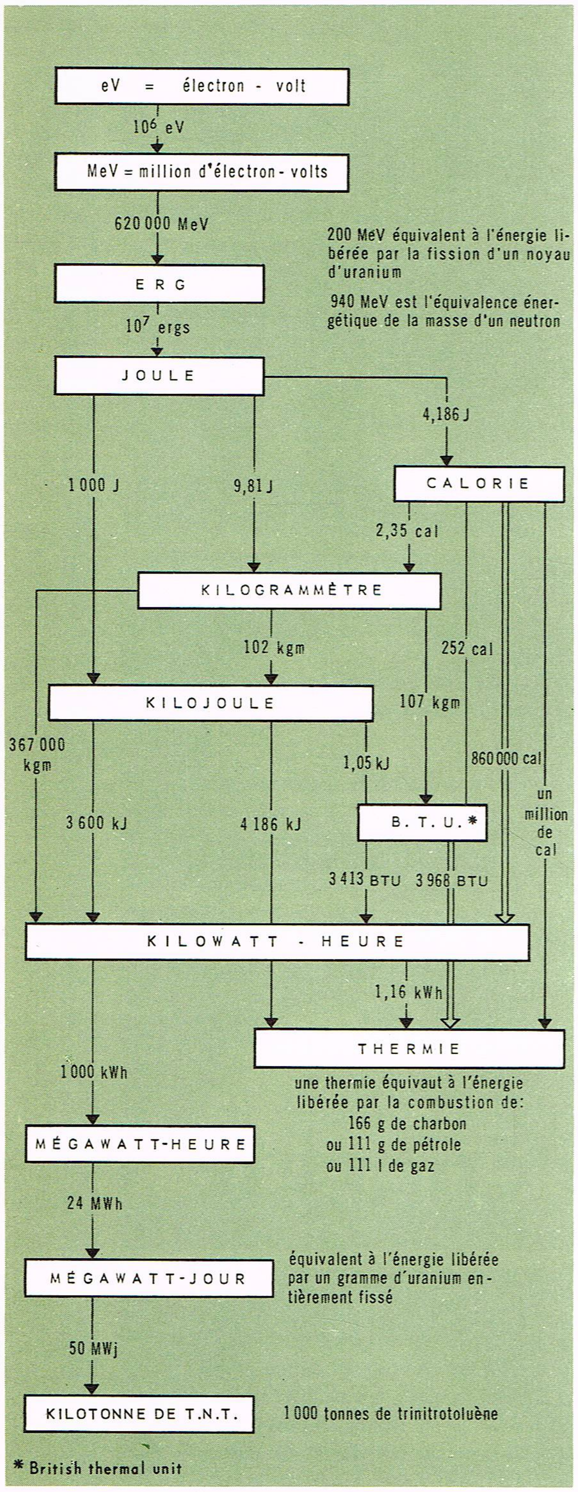
Mais, d'abord, qu'est-ce que l'énergie tout court?

Etymologiquement, *énergie* (εν εργον) signifie « pouvoir faire quelque chose, avoir la possibilité d'effectuer un certain travail ». Pour le physicien, un système, c'est-à-dire un ensemble de corps, possédera de l'énergie s'il peut plus ou moins spontanément évoluer et libérer « quelque chose » qui pourra être transformé en travail. On en peut fournir de nombreux exemples : le charbon a le pouvoir, mélangé dans certaines conditions avec de l'air ou de l'oxygène, de fournir de la chaleur; certains systèmes chimiques appelés piles électriques ont le pouvoir de fournir du courant électrique; quand une masse est animée d'une certaine vitesse, on dit qu'elle possède de l'énergie cinétique, qui peut être transformée, par des machines simples, en travail utile; plus merveilleuse encore est l'énergie qui se manifeste dans les phénomènes de la vie, animale ou végétale : excitantes propriétés des « enzymes » et des acides nucléiques, des cellules et de leur système organisateur. Or, ces diverses sortes d'énergie ne peuvent être extraites que du capital fourni par l'univers, c'est-à-dire par le système solaire : chaleur interne du globe terrestre, énergie cinétique due à sa rotation, rayonnements issus du soleil et déversés en permanence sur la surface de la Terre, la matière enfin dont est constitué le globe terrestre.

Les rayonnements solaires jouent un rôle primordial; sans eux, la vie ne serait pas possible; ils sont responsables de l'énergie que nous recueillons avec les chutes d'eau et les moulins à vent. Par photosynthèse, ils permettent aux plantes de fabriquer de la vie; le même phénomène est probablement à l'origine de la constitution au cours des millénaires d'un stock de charbon, de pétrole et de gaz naturels que nous utilisons abondamment.

énergie classique et nucléaire

Tous ces phénomènes ont un point commun : au cours des transformations énergétiques, que nous venons rapidement de rappeler, les noyaux des différents corps ne subissent aucune modification : l'échafaudage de neutrons et de protons reste



inchangé, les forces nucléaires et électrostatiques qui les maintiennent en équilibre ne fournissent aucun travail.

Si nous examinons la combustion du carbone dans l'oxygène, $C + O_2 \rightarrow CO_2 + 94\,000$ calories, les noyaux de carbone et d'oxygène se retrouvent intacts dans la molécule de CO_2; seuls certains électrons périphériques ont modifié leurs trajectoires pour lier entre eux un atome de carbone et deux atomes d'oxygène, afin de former une molécule de gaz carbonique.

Jusqu'à la fin du siècle dernier, les noyaux des atomes étaient considérés comme insécables, mais en 1896 ce postulat commença à être mis en doute à la suite de la découverte de la radio-activité par Becquerel. Puis, en 1919, Rutherford effectua de véritables transmutations artificielles : en bombardant de l'azote avec des noyaux d'hélium, il obtint de l'oxygène et de l'hydrogène, selon une réaction que l'on peut écrire : ${}^4_2\text{He} + {}^{14}_7\text{N} \rightarrow {}^{17}_8\text{O} + {}^1_1\text{H}$.

Le mécanisme intime d'une telle réaction diffère profondément de celui que nous avons précédemment décrit. Ici, tandis que les électrons ne jouent pratiquement aucun rôle, les noyaux sont modifiés; de plus, au point de vue énergétique, il y a une très grande différence entre ces deux types de réactions : tandis que les premières ne libèrent qu'une énergie de l'ordre de l'électron-volt (que l'on écrit eV) par nucléon mis en jeu, les réactions nucléaires peuvent libérer une énergie de l'ordre du million d'électrons-volts (ou MeV). On appellera énergie atomique ou nucléaire celle que libèrent les réactions modifiant la structure des noyaux.

une découverte d'Einstein

Vers 1905 Einstein arriva à concevoir que la matière de n'importe quel corps renfermait de l'énergie et que, en fait, masse, matière ou énergie étaient à peu près la même chose. La masse est l'une des formes de l'énergie au même titre que l'énergie cinétique, l'énergie potentielle (ou « calorifique ») l'énergie électrique... et il y a une équivalence masse-énergie analogue à l'équivalence chaleur-travail.

Si la masse d'un corps diminue, de l'énergie apparaît sous une autre forme; inversement, si la masse d'un corps augmente, c'est que ce corps a prélevé de l'énergie quelque part dans l'univers.

Si pendant plus d'un siècle, de Lavoisier à Einstein, le monde scientifique a cru à l'invariance de la masse, c'est que l'énergie mise en jeu dans les réactions chimiques ordinaires (celles qui ne sont pas nucléaires) équivaut à des variations de masse impossibles à mesurer. L'application de la formule d'Einstein montre qu'un kilogramme d'une matière quelconque « équivaut » à 25 milliards de kilowatts-heures; il suffirait donc de transformer quelques kilogrammes de matière pour alimenter pendant un an la France entière en énergie électrique.

Devant de tels chiffres, l'énergie des marées et l'énergie que déverse le Soleil à la surface de la Terre deviennent négligeables; la matière est l'ultime refuge où l'énergie s'est accumulée.

quelques réactions faciles à imaginer...

Mais comment faut-il traiter la matière pour qu'elle cède au moins une partie du bien précieux qu'elle contient?

Imaginons d'abord l'opération fictive suivante : prenons deux neutrons et deux protons qui, « pesés » séparément, représentent une masse totale de 4,034 24 u. m. a. (1).

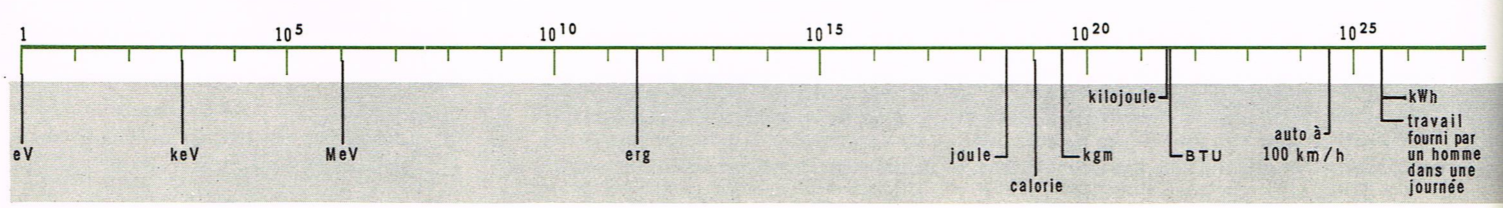
Rassemblons-les pour former un atome d'hélium, l'expérience montre que cet

(1) u. m. a. (unité de masse atomique) vaut $1{,}659\,9.10^{-24}$ g. Ce nombre a été choisi pour que la masse des noyaux de l'atome-gramme d'oxygène soit exactement de 16 g. L'équivalence énergétique d'une masse d'une u. m. a. est

$$\begin{aligned} W = mc^2 &= 1{,}659.10^{-24}\ (3.10^{10})^2 \text{ c. g. s.} \\ &= 1{,}493.10^{-3} \text{ c. g. s.} \\ &= 1{,}493.10^{-3} \times 0{,}62.10^{6} \\ &= 931{,}162 \text{ MeV.} \end{aligned}$$

La masse du neutron vaut 1,008 98 u. m. a. ; celle du proton, 1,008 14 u. m. a.

Valeurs en « électron-volts » d'un certain nombre d'énergies de référence pointées sur une échelle logarithmique.



atome ne pèse que 4,003 87 u. m. a. ; la différence de masse s'est transformée en une énergie de :
[4,034 24 – 4,003 87] · 931,162 = 28,28 MeV.

Il en est toujours de même : quand on forme le noyau d'un corps quelconque à partir de particules primaires telles que neutrons et protons, il y a toujours diminution de masse, c'est-à-dire libération d'énergie. Le quotient de cette énergie par le nombre de nucléons (quatre dans le cas de l'hélium) s'appelle « défaut de masse » ou « énergie de liaison » par nucléon, soit pour l'hélium : 28/4 = 7 MeV. Pour chaque corps elle est indiquée dans le graphique.

Quand on forme un noyau à partir d'un ou plusieurs corps plus simples — c'est-à-dire contenant moins de nucléons —, l'énergie produite est égale à celle qui serait libérée si le corps était formé à partir des nucléons, diminuée de celle qui serait fournie lors de la formation de tous les corps plus simples utilisés (principe du bilan-nucléaire).

Il est facile d'imaginer des réactions nucléaires et, à partir des données numériques du graphique, de calculer l'énergie libérée.

En étudiant toutes les possibilités, on en trouve deux qui sont beaucoup plus intéressantes que toutes les autres :

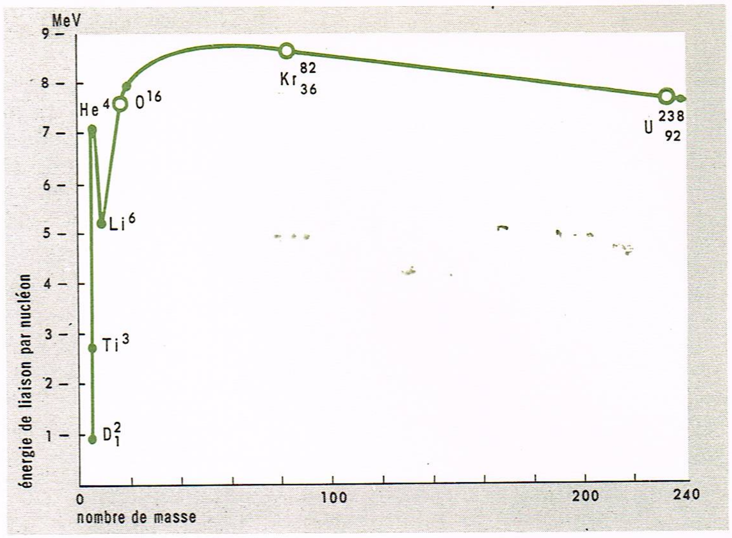
— l'une est de rassembler deux noyaux très légers — par exemple deux noyaux de deutérium — pour en former un moins léger : c'est la fusion;

— l'autre est de « casser » un noyau très lourd en deux morceaux à peu près de même masse : c'est la fission.

Dans les deux cas, on transforme en énergie uniquement les « défauts de masse », qui représentent un peu moins du millième de la masse totale.

L'idéal serait de pouvoir parvenir à une dématérialisation totale. Malheureusement — ou heureusement peut-être —, certaines considérations théoriques laissent supposer qu'on ne peut pas détruire entièrement un neutron ou un proton.

Energie de liaison des noyaux en fonction du nombre de masse.



... mais comment les réaliser ?

Avant d'aller plus loin, il ne faut pas perdre de vue que la règle très simple indiquée ci-dessus pour calculer l'énergie libérée par une réaction nucléaire suppose que **la réaction se produit;** or, une réaction ne se produit que si les corps appelés à réagir sont placés dans certaines conditions favorables. Les seules réactions énergétiquement intéressantes seront donc celles où la réalisation de ces conditions favorables nécessitera une dépense énergétique inférieure à ce qui sera libéré par la réaction.

Finalement, comment pourra-t-on mettre en œuvre cet énorme réservoir d'énergie — même limité aux défauts de masse — que constitue la matière des noyaux? C'est ce que nous allons exposer dans les chapitres suivants en commençant par les propriétés du neutron, particule neutre qui de ce fait peut assez facilement provoquer la fission de certains noyaux lourds, tel l'uranium 235. Les systèmes libérant de l'énergie à partir du phénomène de fission sont appelés des réacteurs nucléaires : ils sont à l'origine de nombreuses réalisations industrielles. Un autre chapitre sera consacré à l'étude de la fusion des noyaux légers.

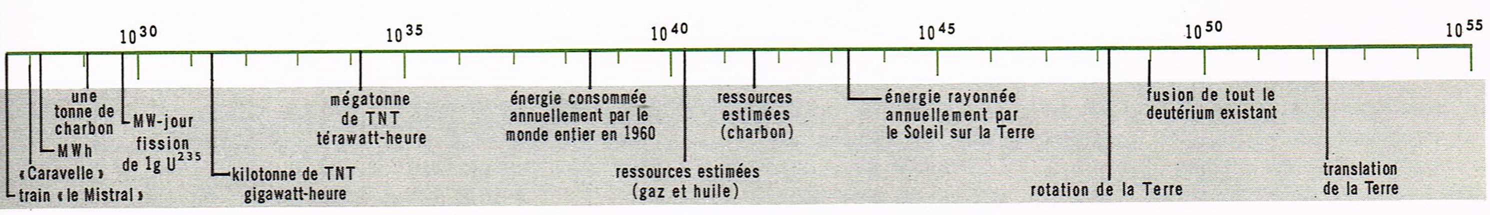
les neutrons, agents de fission

Certains gros noyaux, parmi les plus lourds, sont proches de la limite de stabilité et ne demandent qu'à éclater sous le choc d'une particule. Le neutron est une particule de choix : étant neutre, il n'est pas freiné par les charges des électrons qui entourent le noyau. Effectivement, dès 1938, quatre physiciens allemands : Hahn, Strassmann, Frisch et Lise Meitner, découvrirent et expliquèrent la fission des noyaux d'uranium.

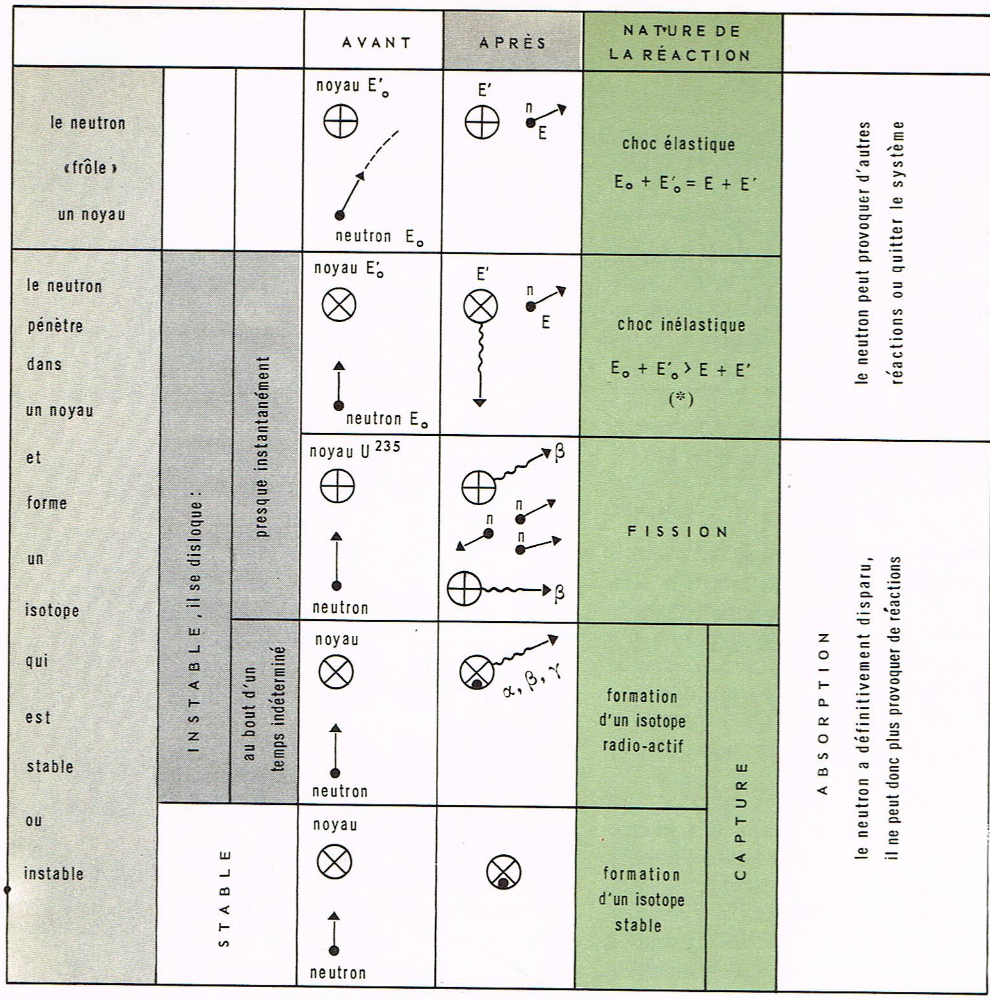
Heurté par un neutron, le noyau d'uranium 235 peut être « fissé », scindé en deux morceaux principaux ayant des masses voisines : fait remarquable (découvert par Joliot en 1939), le noyau fissé émet en sus des deux corps principaux environ trois neutrons. Ces neutrons peuvent à leur tour provoquer d'autres fissions et on arrive ainsi à l'idée de réaction en chaîne.

Avant de poursuivre l'étude de la fission et d'étudier comment une réaction en chaîne peut s'entretenir, nous allons décrire les autres réactions que les neutrons sont susceptibles de provoquer.

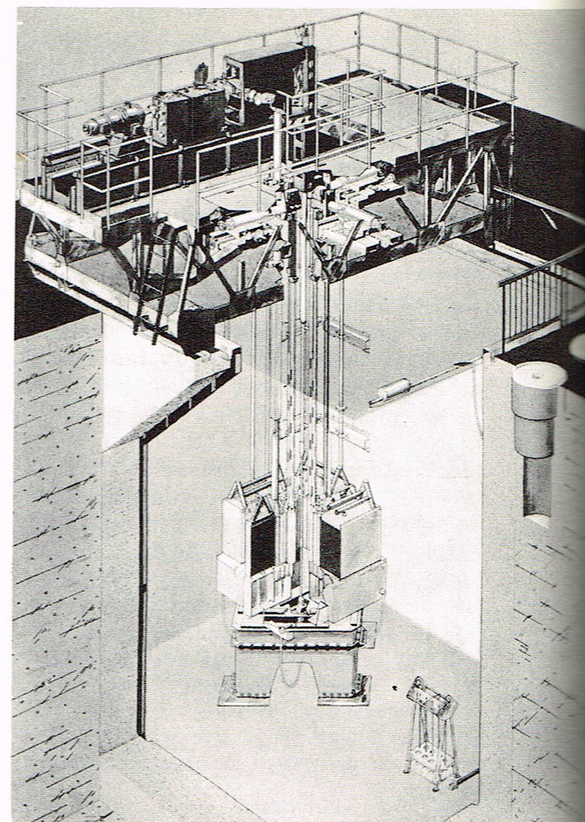
Que peut faire un neutron se déplaçant dans un certain volume de matière?



Réactions d'un neutron avec la matière.



(*) Pour vérifier le principe de la conservation de l'énergie, il faut ajouter au second membre l'énergie du photon γ.



Coupe de la pile « Minerve » du Centre de Fontenay-aux-Roses, spécialement utilisée pour mesurer les sections efficaces et autres caractéristiques neutroniques des matériaux.

Quelles réactions peut-il provoquer sur les noyaux la constituant?

L'expérience montre que pour des conditions initiales parfaitement déterminées, il peut se produire plusieurs phénomènes qui, bien entendu, s'excluent.

A. Le neutron ne rencontre aucun noyau, il sort du volume considéré sans provoquer aucune réaction.

B. Le neutron « frôle » un noyau, et se « réfléchit » sur lui; il y a un choc type « boule de billard », qui se produit sans perte d'énergie cinétique; ce neutron peut ensuite soit quitter le système, soit subir un des sorts décrits ci-après.

C. Le neutron pénètre dans le noyau, et commence donc par former un isotope du noyau initial. Il peut alors se présenter plusieurs cas :

1. L'isotope ainsi formé est stable;

2. L'isotope formé est radio-actif — il se désintègre en émettant suivant les cas des rayonnements α, β, γ ou neutrons. L'ensemble des phénomènes 1 et 2 ci-dessus s'appelle la « capture »;

3. L'isotope se disloque instantanément en un temps de l'ordre de 10^{-14} seconde — et cette dislocation peut s'effectuer suivant deux processus principaux :

a) il y a réémission d'un neutron. Le noyau est donc revenu à son état initial, mais il y a une perte d'énergie cinétique; cette perte est compensée par l'émission d'un rayonnement γ — on parle alors de « choc inélastique »;

b) il y a fission, c'est-à-dire que le noyau initial se coupe en deux morceaux principaux ayant des nombres de masse voisins et émet en même temps quelques neutrons.

probabilités et sections efficaces

Nous voyons ici un des phénomènes caractéristiques de la nouvelle physique. A partir de conditions initiales bien déterminées (compte tenu de nos moyens d'investigation), on ne peut pas prédire sûrement la réaction qui va se produire. **L'évolution d'une particule ou d'un événement est aléatoire.**

Mais dès que nous considérons non un seul neutron mais plusieurs, le caractère aléatoire disparaît et on peut affirmer que chaque événement possible se produira avec une certaine fréquence : **on ne peut plus parler de certitude, mais de probabilité ou de fréquence.**

Pour chiffrer ces probabilités, ou encore le nombre de chances qu'un neutron a de provoquer telle ou telle réaction, on a recours à l'expérience et il est commode d'introduire la notion de « section efficace ».

Considérons un cube d'un centimètre de côté contenant N noyaux identiques qui ne peuvent provoquer qu'une seule réaction : la capture, par exemple.

Ce cube considéré comme une cible est bombardé par un faisceau de neutrons, ayant tous la même vitesse v, perpendiculaire à une face du cube. Provenant d'une source très éloignée, ce

Phot. C. E. A.

faisceau est supposé homogène et contient n neutrons par centimètre cube.

Il est évident que les neutrons qui pénétreront dans le cube pendant une seconde sont ceux dont la distance au cube est inférieure à v; ils sont donc contenus dans un parallélépipède de hauteur v ayant comme base 1 cm²; leur nombre est égal à nv, quantité qui est en général désignée par φ et appelée : « flux neutronique ». Les valeurs courantes de ce flux sont de l'ordre de 10^{13}-10^{14} n/cm²/sec.

Finalement, on a un cube qui contient N noyaux qui est traversé pendant une seconde par nv projectiles; on peut toujours écrire que le nombre de réactions qui se produisent en une seconde dans notre cube de 1 cm³ est égal à :

$$nvN\sigma = \varphi N\sigma.$$

La quantité σ ainsi définie s'appelle « section efficace de capture ». Le succès de cette notion de section efficace est dû au fait que l'expérience montre que *σ ne dépend que de la nature des noyaux constituant la cible et de la vitesse des neutrons incidents, ou encore de leur énergie* $\frac{1}{2}mv^2$.

σ est de même nature qu'une surface; on l'exprime en général en barn, unité très petite (10^{-24} cm²).

En général, un noyau peut subir plusieurs types de réactions. On définit alors de manière analogue une section efficace pour chaque type : capture, fission, choc élastique... et une section totale qui est la somme des sections particulières.

Si le cube que nous avons considéré contenait des noyaux d'uranium 235, les neutrons pourraient provoquer trois sortes de réactions : fission, capture ou choc élastique, dont les sections efficaces valent approximativement : 600, 100 et 10 barns pour les neutrons ayant l'énergie thermique 1/40 eV; cela signifie entre autres que sur un grand nombre de réactions, chaque fois qu'un neutron subit un choc élastique, 10 sont capturés et 60 produisent des fissions créant ainsi 150 neutrons environ; et il faut bien noter qu'aucun moyen connu ne permet de favoriser une réaction aux dépens des autres.

De même si on a un milieu hétérogène contenant des noyaux de plusieurs espèces différentes, le nombre de réactions par centimètre cube est égal à la somme des réactions qui se produiraient si chaque espèce de noyaux était seule.

Prolongeant les arêtes de notre cube qui sont parallèles à la vitesse des neutrons, nous définissons ainsi un parallélépipède rectangle ayant une section droite carrée de 1 cm de côté et nous supposons qu'à l'intérieur de ce volume la densité de noyaux cibles est constante et égale à N.

Au cours de chaque seconde φ_0 neutrons arrivent sur la face d'entrée A_0 du parallélépipède. On peut mesurer le nombre φ_l de neutrons qui pendant le même temps traversent la section de 1 cm² située à la distance l. Ce nombre est inférieur à φ_0 et la différence est égale au nombre de neutrons qui ont été capturés entre les deux faces considérées. φ_l diminue et tend vers 0 quand l augmente.

Le calcul montre que la moyenne des chemins parcourus par un grand nombre de neutrons, entre la face d'entrée et le point où ils sont capturés, est égale à $\lambda_c = 1/N\sigma_c$.

λ_c est dit « libre parcours moyen de capture ». On définit de même un « libre parcours moyen de fission ».

Les tables fournissant les sections efficaces permettent de calculer le libre parcours de fission dans l'uranium 238 (29 cm pour des neutrons de 10 MeV) et dans l'uranium 235 (0,3 mm pour des neutrons d'énergie thermique, c'est-à-dire de 0,025 eV).

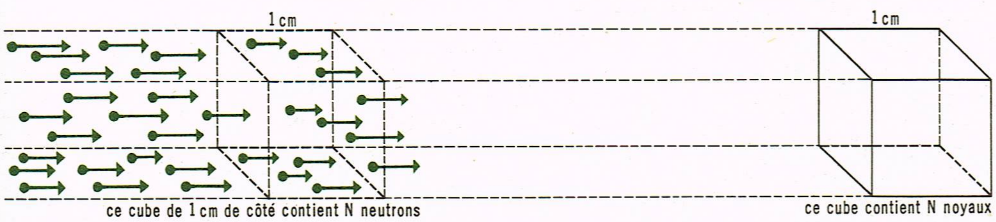
V. tableau p. 252.

Quand on se place aux faibles énergies il n'y a que trois noyaux fissiles : uranium 233, uranium 235 et plutonium 239 (quand on parle de noyaux fissiles sans plus préciser, on sous-entend toujours : aux faibles énergies de l'ordre de l'électron-volt). Si l'on passe aux énergies élevées (de l'ordre du méga-électron-volt), il faut ajouter deux autres noyaux : l'uranium 238 et le thorium 232.

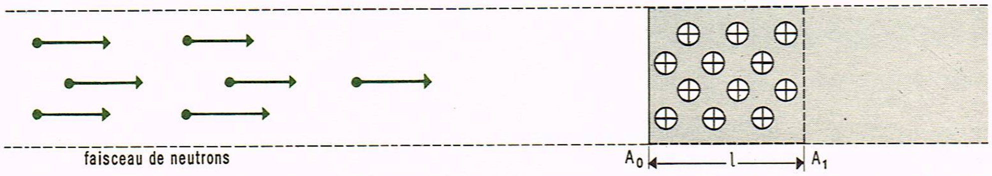
Le plutonium 239 n'existe pas dans la nature, c'est un radio-isotope artificiel fabriqué, comme nous le verrons plus loin, dans les piles atomiques. Il est surtout utilisé jusqu'ici pour fabriquer les bombes atomiques.

L'uranium 233 n'existe pas non plus dans la nature. L'uranium 235 existe dans la proportion de 7 pour 1 000 dans l'uranium naturel; le complément est de l'uranium 238.

Il est intéressant de comparer les sections efficaces aux dimensions des noyaux et des atomes; un noyau lourd présente une surface apparente de l'ordre de 10^{-24} cm², c'est-à-dire de 1 barn.



Chaque neutron de faisceau incident « voit » le cube de matière comme un carré de 1 cm de côté et les noyaux comme N disques ayant chacun une surface égale à σ. Donc, sur une surface de 1 cm, on a N cibles qui occupent une surface Nσ.
La probabilité qu'a un neutron de rencontrer un noyau-cible est égale au rapport de ces deux surfaces : $\frac{N\sigma}{1} = N\sigma$.
φ neutrons auront une probabilité φ fois plus importante et provoqueront φNσ réactions.



Le nombre de neutrons φl qui, par seconde, traverse la section A_1 est inférieur à celui qui entre dans la matière par la face A_0.
φ_l diminue quand la distance l augmente et tend vers 0.
La moyenne pondérée de l est le libre parcours moyen des neutrons dans la matière; il est égal à $l = \frac{1}{N\sigma}$.

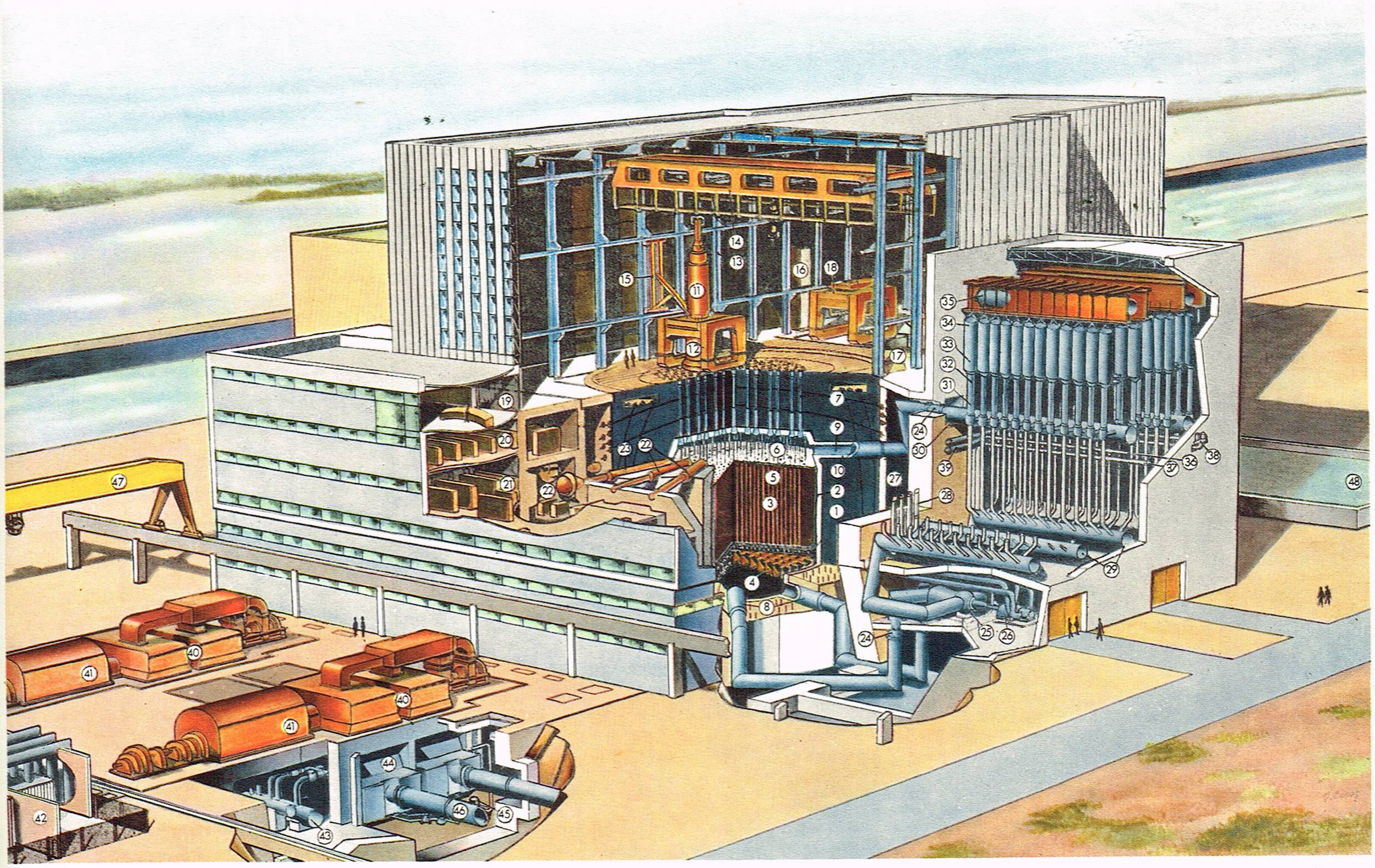
qu'est-ce qu'une fission ?

Pour l'uranium 235, le processus est le suivant : un neutron est d'abord capté et forme l'isotope 236, qui, en un temps très court (10^{-14} seconde environ), explose en donnant deux noyaux de masse comparable et ν neutrons.

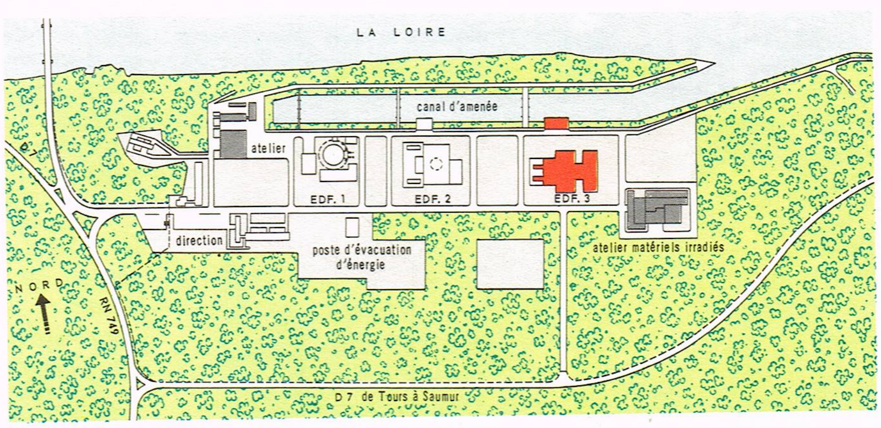
$$U^{235}_{92} + n^1_0 \rightarrow U^{236}_{92} \rightarrow A + B + \nu n^1_0 + \text{énergie.}$$



La centrale nucléaire de Chinon dispose de trois réacteurs : E.D.F. 1 (bâtiment sphérique), E.D.F. 2 (au milieu) E.D.F. 3 (à l'arrière-plan, en construction). *Phot. H. Barangé.*

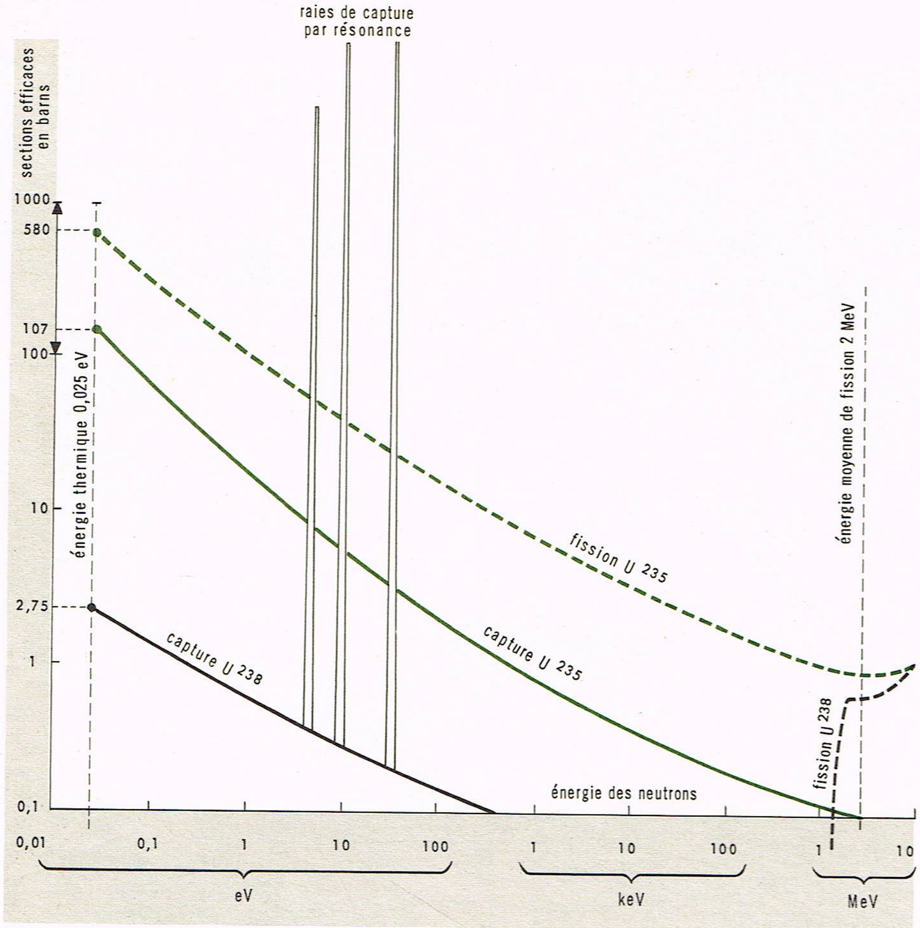


Ecorché de la centrale E.D.F. 3.



Plan de la centrale nucléaire de Chinon.

1 Caisson en béton précontraint du réacteur.
2 Membrane d'étanchéité.
3 Noyau actif de graphite.
4 Aire de support.
5 Pavés de lestage.
6 Tube guide de barre de contrôle.
7 Puits de barre de contrôle.
8 Entrée du CO_2.
9 Sortie du CO_2.
10 Calorifuge.
11 Machine principale.
12 Pièce intermédiaire.
13 Château.
14 Pont de manutention du château.
15 Chariot de raccordement électrique et pneumatique.
16 Pondoir.
17 Transbordeur.
18 Chariot de manutention de la pièce intermédiaire.
19 Salle de commande.
20 Salle des armoires d'électronique.
21 Relayage.
22 Sortie des tubes D. R. G.
23 Thermocouples.
24 Obturateur.
25 Soufflantes.
26 Turbine auxiliaire.
27 Vapeur surchauffée vers la turbine auxiliaire.
28 Retour de la vapeur des turbines auxiliaires.
29 Sortie du CO_2 de l'échangeur.
30 Entrée du CO_2 dans l'échangeur.
31 Resurchauffeur.
32 Surchauffeur.
33 Economiseur.
34 Vaporisateur.
35 Ballon.
36 Collecteur de la vapeur resurchauffée.
37 Collecteur de la vapeur surchauffée.
38 Soupapes.
39 Collecteur de la vapeur resurchauffée vers la turbine principale.
40 Turbines principales.
41 Alternateur.
42 Transformateur.
43 Poste de réchauffage.
44 Condensateur.
45 Pompes d'extraction.
46 Tuyauteries Bonna.
47 Portique.
48 Piscine de désactivation.



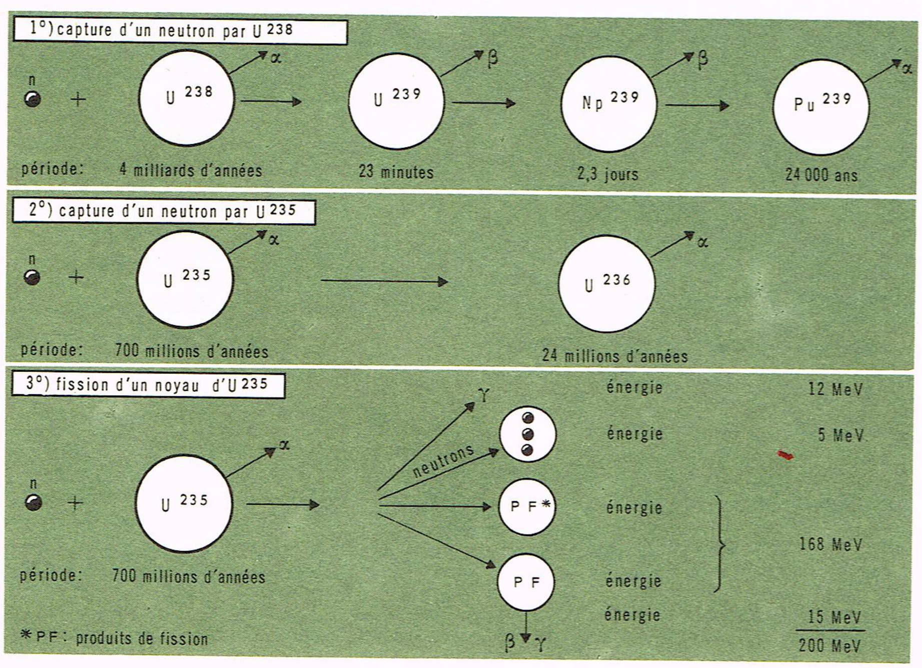
Sections efficaces de capture et de fission pour U^{235} et U^{238}.

U^{235} est fissile sans solution de continuité, depuis 10 MeV jusqu'à l'énergie thermique.

U^{238} n'est fissile que pour les énergies supérieures à 1 MeV, et la section efficace est faible, inférieure à un barn.

La section de capture de l'U^{235} est beaucoup plus importante que celle de l'U^{238}, abstraction faite des raies de résonance.

Absorption des neutrons.
Réactions possibles d'un neutron thermique dans l'uranium naturel.



En fait, il y a plusieurs types de fissions, qui diffèrent l'une de l'autre par la valeur du nombre ν de neutrons émis, par leur énergie, et par la nature des corps A et B qui sont dits « produits de fission ».

Si nous examinons individuellement différents types de fissions, ν peut varier de 0 à 6; mais si l'on considère l'ensemble d'un grand nombre de fissions, ν a une valeur moyenne bien déterminée : 2,5 pour les noyaux d'uranium et 2,9 pour ceux du plutonium.

ν est le nombre de neutrons produits en moyenne par noyau fissé; mais ces noyaux sont aussi capturants : si on rapporte le nombre de neutrons créés au nombre de neutrons absorbés (total de ceux qui provoquent la fission et de ceux qui sont absorbés), on définit un paramètre η qui est inférieur à ν. Des considérations simples montrent que $\eta = \frac{\nu}{1 + \sigma_e/\sigma_f}$.

En général, ce paramètre caractérise mieux que ν l'efficacité d'un noyau fissile. Pour les neutrons thermiques (0,025 eV), η est voisin de 2, tant pour l'uranium que pour le plutonium. Par contre, avec les neutrons rapides : $\eta = 2{,}24$ pour l'uranium et 2,74 pour le plutonium.

Les neutrons émis n'ont pas tous la même énergie; quelques-uns ont une vitesse très faible (énergie presque nulle), d'autres se déplacent au cinquième de la vitesse de la lumière (énergie environ 15 MeV); l'énergie moyenne est de 2 MeV (ce qui correspond à une vitesse de 20 000 km/s).

Les deux noyaux, appelés « produits de fission », ont des masses comparables et possèdent une énergie cinétique très importante, voisine de l'énergie libérée par la réaction de fission, soit environ 200 MeV. Cette énergie, vue à l'échelle microscopique, est entièrement cinétique.

Il existe une quarantaine de produits de fission — plus de 100 si l'on tient compte des isotopes — dont les masses atomiques varient de 75 à 160; chacun d'eux est créé dans une proportion bien déterminée.

Presque tous sont radio-actifs et se désintègrent en émettant des radiations α, β ou γ. Quelques-uns possèdent un mode de désintégration inconnu des radio-isotopes naturels; ils émettent des neutrons. Cette émission a lieu plusieurs dizaines de secondes après la fission. Ce sont les « neutrons retardés »; bien qu'en proportion très faible — moins de 1 p. 100 des neutrons émis au moment de la fission (appelés souvent « neutrons prompts ») —,

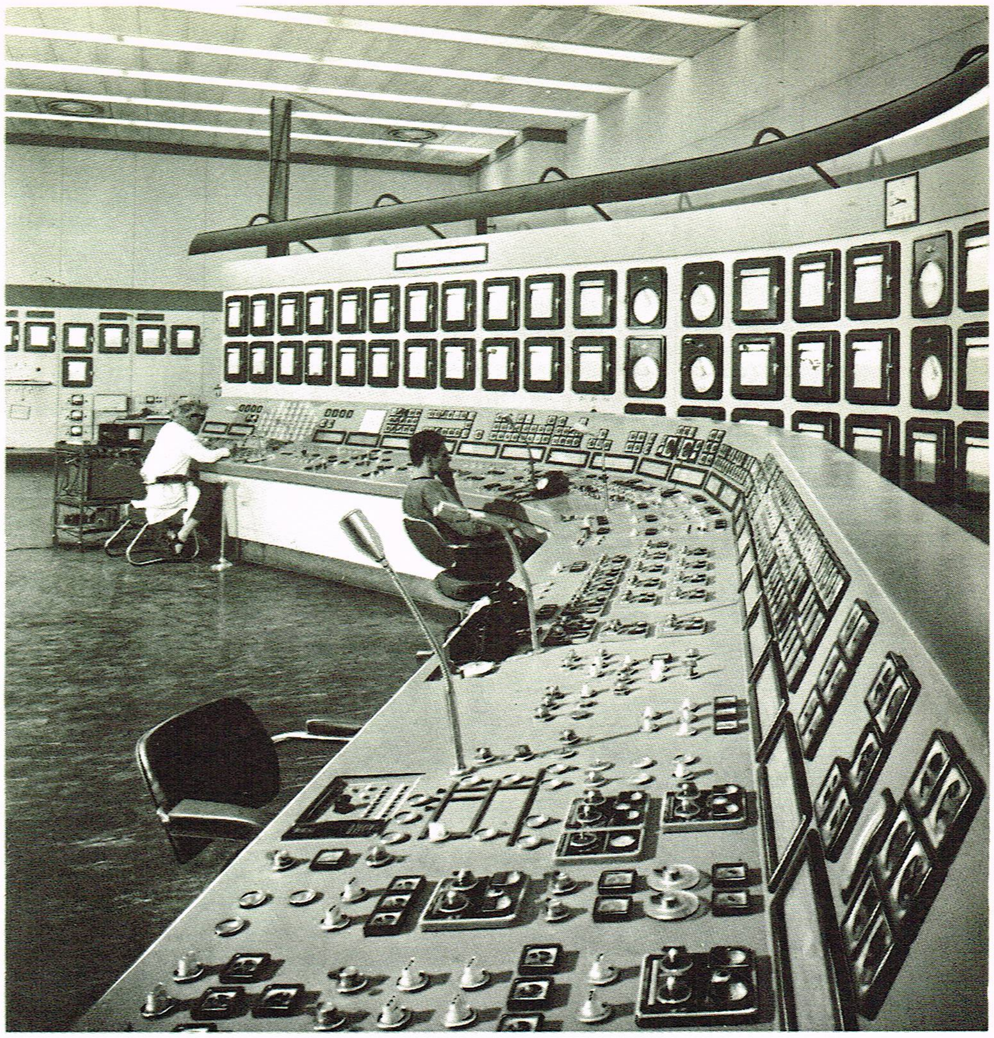
ils jouent un rôle capital dans le contrôle des réacteurs nucléaires. Sans eux, l'homme n'aurait pas pu construire de piles atomiques, les applications de la nouvelle forme d'énergie que nous étudions auraient été limitées aux bombes.

Finalement l'intérêt de la fission résulte des trois caractéristiques suivantes :

— l'importance de l'énergie libérée rapportée aux masses mises en jeu — près de 1 MeV par nucléon. Cette énergie se retrouve presque totalement dans l'énergie cinétique des produits de fission; comme le phénomène est isotrope, il se produit une dégradation, et on a finalement de la chaleur;

— la possibilité de réaliser des réactions en chaîne, puisque la réaction qui consomme un neutron en produit en moyenne 2,5;

— l'existence des neutrons retardés.



Le pupitre de commande d'un réacteur de puissance (G2 à Marcoule). En premier lieu, il s'agit de contrôler le flux de neutrons thermiques.

pour que diverge un réacteur

D'après les caractéristiques de la fission, il n'est pas *a priori* impossible d'obtenir une succession ininterrompue de réactions, chacune d'elles, grâce aux neutrons émis, pouvant en provoquer une ou plusieurs autres. Mais on peut se demander si, pratiquement, de telles réactions « en chaîne » sont facilement réalisables et s'il est possible qu'elles s'auto-entretiennent un temps suffisamment long.

Il ne faut pas perdre de vue qu'un neutron ne provoque pas à coup sûr une fission; même dans un matériau fissile pur, chaque neutron peut, en effet, quitter le système sans provoquer de réaction ou être capturé sans provoquer de fission. Et nous comprenons sans peine que la réaction ne sera auto-entretenue que si elle produit plus de neutrons qu'il n'en disparaît par captures, fissions et fuites.

Pour préciser, considérons un ensemble d'éléments de matière, pouvant appartenir à des espèces physiques ou chimiques différentes (mais stables) — l'une d'elles au moins étant fissile —, répartis géométriquement d'une manière quelconque, mais bien déterminée. Ce système est donc en équilibre et peut y demeurer. Bien entendu, il s'agit là d'un cas très théorique, car il y a toujours des neutrons parasites — provenant, par exemple, des rayons cosmiques — qu'on ne peut empêcher et qui vont provoquer des fissions dans le système.

Supposons maintenant que l'on introduise brusquement dans le système N neutrons. Au bout d'un certain temps, ces N neutrons, qui constituent la première génération, auront disparu, pour des causes diverses que l'on peut classer en trois groupes :

1. N_1 auront provoqué une fission;
2. N_2 auront été capturés par un noyau quelconque du système;
3. N_3 s'en seront enfuis, c'est-à-dire qu'ils auront fui à travers la surface limitant le système, sans espoir de retour.

Comme nous raisonnons toujours sur un grand nombre de particules, on peut définir pour chacune des trois catégories une durée de vie moyenne; afin de simplifier, nous supposerons que celle qui est relative à la fission est supérieure aux deux autres.

L'effectif de la deuxième génération sera égal à νN_1, qu'on peut écrire $N(\nu N_1/N)$; ces neutrons se trouvant dans une situation « quasi » identique à ceux de la première, la même proportion produira des fissions et l'effectif de la troisième génération sera $N(\nu N_1/N)^2$; on verrait de même que celui de la $(p+1)^{ème}$ serait $N(\nu N_1/N)^p$.

les systèmes critiques

Si νN_1 est supérieur à N, chaque génération aura un effectif supérieur à celui de la génération précédente; on dira alors qu'on a des réactions en chaîne auto-entretenues, ou un système sur-critique, ou encore une chaîne divergente.

Si les disparitions sont supérieures aux créations ($N > \nu N_1$) on aura un système sous-critique (chaîne convergente); c'est-à-dire qu'après un nombre de générations plus ou moins important, toutes les réactions seront pratiquement arrêtées.

Dans le cas limite où les créations sont exactement égales aux disparitions :

$$(\nu N_1 = N),$$

on dit que le système est juste critique.

Phot. Jahan (C. E. A.).

Le rapport $\nu N_1/N$, qui est une caractéristique essentielle du système, s'appelle le facteur de multiplication effectif, on le désigne par k_{eff}, ou simplement par k quand il n'y a pas de confusion possible.

La quantité $(k\text{-}1/k)$ s'appelle la « réactivité du système »; c'est un nombre qui est toujours petit, sauf dans le cas où le système considéré est une bombe; on l'exprime en général en cent-millièmes (k-1/k diffère peu alors de k-1, puisque k est voisin de 1).

Si un système a, par exemple, une réactivité de 0,034, on dira que ce paramètre est égal à 3 400 cent-millièmes, ou encore : 3 400 p. c. m., cette abréviation signifiant : parties pour cent mille.

Donc, un système divergera ou convergera suivant que sa réactivité sera positive ou négative.

Il faut bien voir que la valeur de k_{eff} est uniquement fonction des caractéristiques physiques — sections efficaces entre autres — des noyaux qui constituent ce système, et de leurs positions relatives, et est indépendante de l'effectif des générations.

La condition de divergence : $k_{eff} \geqslant 1$ peut s'écrire

$$(\nu N_1 \geqslant N = N_1 + N_2 + N_3)$$

ou encore

$$(\nu - 1)\ N_1 - N_2 \geqslant N_3.$$

Fissions et captures se produisant dans chaque élément de volume, les termes N_1 et N_2 sont proportionnels au volume du système, c'est-à-dire au cube d'une dimension caractéristique. (La notion de dimension caractéristique ne présente aucune difficulté quand la géométrie du système est très simple : sphère ou cube. « L » sera, par exemple, le diamètre de la sphère; on voit facilement que si on double le diamètre d'une sphère, son volume est multiplié par $2^3 = 8$, tandis que sa surface est multipliée par $2^2 = 4$. Cette dimension caractéristique est beaucoup plus difficile à préciser quand la géométrie est compliquée.) Le premier membre de l'inégalité peut donc s'écrire

$$(\nu - 1)\ N_1 - N_2 = aL^3.$$

Les fuites ne se produisant que par les éléments de surface extérieurs, N_3 est proportionnel à la surface de l'enveloppe, c'est-à-dire au carré de la même dimension caractéristique et le second membre peut s'écrire $N_3 = bL^2$.

On a finalement $aL^3 \geqslant bL^2$

$$\text{ou } L \geqslant \frac{b}{a}.$$

Il ne peut y avoir auto-entretien que si une dimension principale est supérieure à un certain minimum. De ces considérations très simples, on déduit les notions de masse et de volume critiques.

La terminologie « masse critique » ne signifie pas grand-chose si l'on ne définit pas en même temps dans quel volume cette masse est répartie. En général, quand on parle de masse critique sans autre précision, on sous-entend que la matière fissile se présente sous la forme d'une sphère pleine, et la « masse critique » est alors la masse d'une telle sphère qui constitue un système juste critique ($k_{eff} = 1$). Mais il est bien évident que si la même masse de matière fissile a la forme d'un cylindre très allongé, le système sera sous-critique ($k_{eff} < 1$), car les fuites de neutrons seront relativement beaucoup plus importantes que dans le cas de la sphère.

Il se trouve que les masses critiques — en sphères pleines — des matières fissiles sont toutes de l'ordre de quelques kilogrammes, mais c'est là un pur hasard. Si ces masses avaient été mille fois plus grandes, il n'aurait probablement jamais été question d'énergie nucléaire, et si elles étaient mille fois plus petites on pourrait concevoir des réacteurs qui fonctionneraient d'une manière très différente des réacteurs actuels.

La formule définissant k_{eff} peut s'écrire

$$k_{eff} = \frac{\nu N_1}{N_1 + N_2 + N_3} = \frac{\nu N_1}{N_1 + N_2} \cdot \frac{N_1 + N_2}{N_1 + N_2 + N_3}$$

La première fraction du dernier membre est la valeur qu'aurait k_{eff} en l'absence de fuites, c'est-à-dire si le système avait des dimensions infinies. Cette valeur, appelée « k infini », est donc indépendante des dimensions, elle n'est fonction que de la répartition des différents matériaux et de leur nature.

La seconde fraction, toujours inférieure à un, est la proportion de neutrons qui ne fuient pas, c'est la probabilité de « non-fuite »; nous la désignerons par P. Finalement $k_{eff} = k_\infty P$.

P dépend des dimensions du système, il croît en même temps que celles-ci et atteint l'unité quand elles sont très grandes. En général, on cherche à augmenter k_{eff}; on y arrive :

— En diminuant le terme N_3/N_1, c'est-à-dire les fuites relatives. Pour cela, il faut augmenter les dimensions générales du système, ou encore l'entourer d'un « réflecteur », corps qui dévie les trajectoires des neutrons qui fuient de manière qu'une proportion importante soit renvoyée vers le système;

— En diminuant le terme N_2, c'est-à-dire les captures. On ne peut pas agir sur les captures provoquées par les noyaux fissiles; mais on le peut sur les captures parasites en choisissant convenablement les matériaux de structure.

Si dans certains éléments de volume on place des noyaux très capturants, les pertes par capture augmenteront, et il est évident que si on peut facilement enlever et replacer de tels noyaux on aura un moyen commode pour faire varier k_{eff}.

C'est le principe des barres de contrôle et de sécurité dont nous parlerons plus loin;

— En augmentant enfin le nombre de fissions N_1.

Ce nombre est égal — par unité de temps et de volume — à : $\varphi M \sigma_f$.

On peut donc soit augmenter ce nombre M de noyaux fissiles par unité de volume, c'est-à-dire utiliser des combustibles enrichis, soit augmenter la section efficace σ_f, c'est-à-dire diminuer l'énergie des neutrons.

Comme nous l'avons déjà signalé, k_{eff} est indépendant de φ, car toute variation du flux entraîne des variations proportionnelles des trois termes N_1, N_2, N_3.

savoir contrôler une pile atomique

Nous avons indiqué plus haut que les neutrons de la deuxième génération étaient dans une situation « quasi » identique à ceux de la première; ce n'est pas tout à fait exact : les premiers neutrons, en provoquant des fissions, ont modifié la concentration des noyaux fissiles et de nouveaux noyaux (produits de fission) ont fait leur apparition.

Certes, pendant une génération, le nombre de noyaux modifiés est extrêmement faible, mais après une journée de fonctionnement continu le nombre de générations est de l'ordre de cent millions et le cumul de ces petites modifications arrive à produire des changements notables.

Sans entrer dans trop de détails, disons :
— qu'un grand nombre de noyaux créés sont radio-actifs — entre autres les produits de fission —, donc dangereux pour les êtres vivants. On s'en protège par une enveloppe dite « bouclier biologique »;
— que, parmi les produits de fission, le samarium 149 et le xénon 135 ont des

sections de capture très importantes et diminuent sensiblement le facteur k_{eff};
— que les noyaux fissiles sont capturants et qu'une partie de ces noyaux est stérilisée par capture de neutrons;
— que deux noyaux (fertiles), U 238 et Th 232, jouent un rôle très important : par capture ils se transforment en uranium 239 et thorium 233, puis, par désintégration, en deux noyaux fissiles : le plutonium 239 et l'uranium 233.

Le rapport entre le nombre de noyaux fissiles créés et celui qui est détruit s'appelle le « facteur de conversion ».

En général, k_{eff} diminue au fur et à mesure du défilement des générations, et devient inférieur à un; alors le système ne peut plus diverger. Mais nous avons vu que l'on peut faire varier k en agissant sur les barres de contrôle — qui sont constituées par des noyaux très absorbants, cadmium par exemple —; dès que k sera légèrement inférieur à un, on sortira un peu les barres du système afin de diminuer la proportion de neutrons capturés, k deviendra à nouveau supérieur à l'unité, le système divergera à nouveau, puis s'arrêtera; on peut recommencer plusieurs fois cette opération, mais quand les barres seront complètement sorties, le réacteur sera définitivement arrêté, le combustible nucléaire est alors usé, il faut le décharger et le remplacer par une charge neuve.

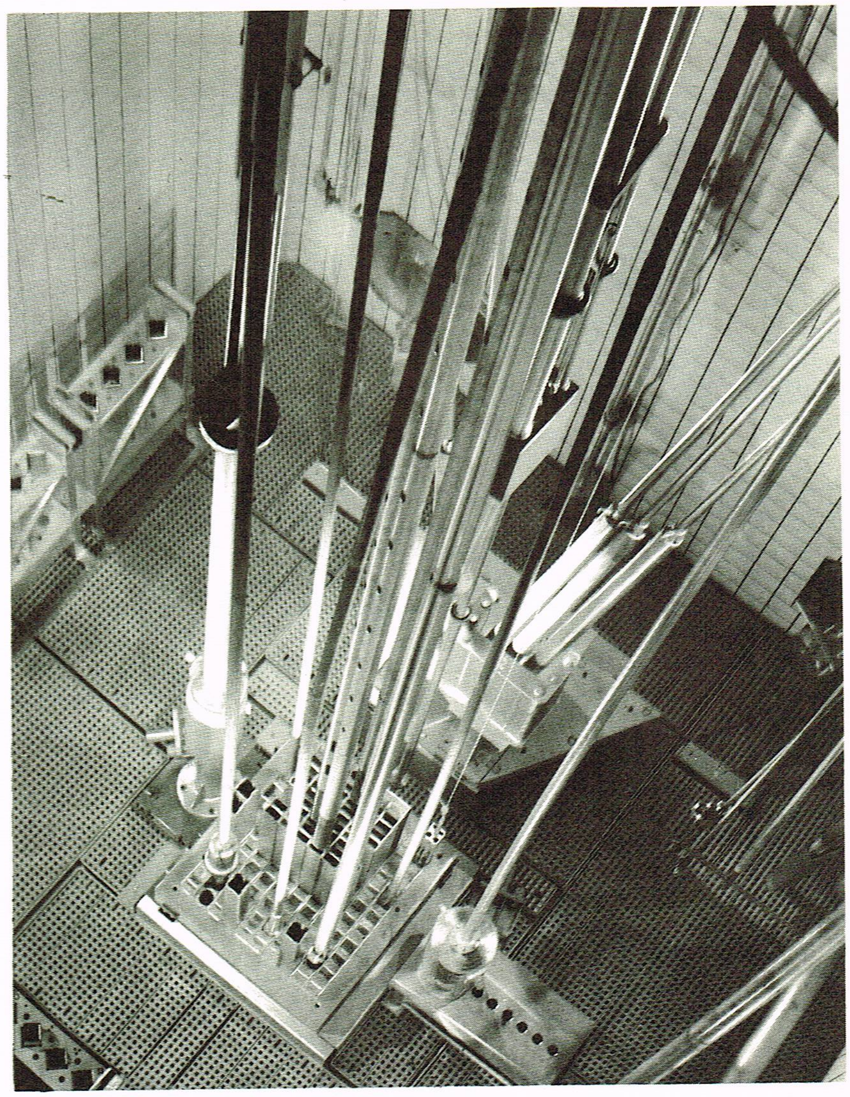
L'énergie qui a été fournie par l'unité de masse du combustible, ou taux de combustion — *burn-up* dans les ouvrages anglo-saxons —, est analogue au pouvoir calorifique des combustibles classiques. Elle n'en diffère que par l'ordre de grandeur (elle s'exprime en mégawatts-jour par tonne de combustible [1]; un mégawatt-jour correspond à peu près à l'énergie libérée par la fission de un gramme de matière) et par le fait que s'il est relativement facile de brûler totalement du charbon ou du pétrole, il est impossible d'arriver à la même performance avec un combustible nucléaire. Il faut toujours avoir dans les réacteurs en fonctionnement une quantité de combustible très supérieure à un minimum fixé par la masse critique. Quand k est supérieur à l'unité, la puissance thermique libérée par les réactions nucléaires augmente avec le temps; si on ne veut pas arriver à la destruction du système par élévation de température, il faudra extraire l'énergie au fur et à mesure de sa production. **Finalement, ce sont les possibilités du système qui extrait la chaleur qui fixent la puissance nominale d'un réacteur nucléaire.**

Dans un grand nombre de réacteurs, les neutrons, dès leur création, sont considérablement ralentis jusqu'à un niveau énergétique égal à : kT, T étant la température absolue du système. Or, au cours de la montée en puissance du réacteur, sa température augmente, il en est de même de l'énergie des neutrons; cela a pour conséquence une diminution des sections efficaces, donc de k_{eff}. La réactivité d'un réacteur diminue quand il s'échauffe.

Il y a là un phénomène de compensation, mais qui, en général, n'a pas une amplitude suffisante pour « contrôler » le système, c'est-à-dire pour maintenir k_{eff} à une valeur très voisine de 1, ce qui est absolument indispensable comme le montrent les considérations ci-après.

La vie d'une génération de neutrons dans les réacteurs thermiques est de l'ordre du millième de seconde; supposons alors que, pour une cause quelconque, la valeur de k passe brusquement de 1 à 1,01, le calcul indique que la puissance serait multipliée par 20 000 au bout d'une seconde : le système serait détruit bien avant que ce chiffre soit atteint. Finalement, on arrive à contrôler les réacteurs grâce aux neutrons retardés, dont l'effectif n'est que 7,5 p. 1 000 du total; on a donc

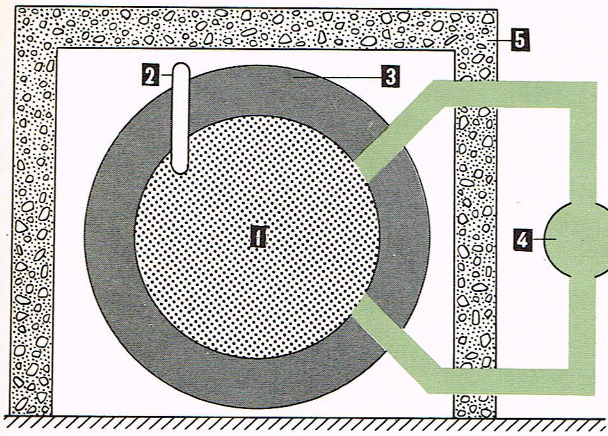
$$k_{eff} \text{ prompt} = \frac{k_{eff} \text{ total}}{1{,}007\,5}.$$



Au Centre d'études nucléaires de Grenoble : vue des barres de contrôle plongeant dans le cœur du réacteur « Mélusine ».

[1] Nous verrons plus loin que la matière fissile, en général mélangée avec des matières non fissiles, est contenue dans une enveloppe étanche appelée « gaine ». Par définition, nous appellerons combustible tout ce qui est à l'intérieur de la gaine.

Phot. C. E. N. G.



1. **Système contenant, entre autres, des noyaux fissiles.**
2. **Barre de contrôle.**
3. **Réflecteur.**
4. **Dispositif pour extraire l'énergie.**
5. **Protection biologique — bouclier.**

Le réglage des barres de contrôle et de sécurité est fait de manière que

$$1 \leqslant k_{eff} \text{ total} < 1{,}007\,50.$$

Le réacteur ne peut pas diverger avec les neutrons prompts puisque (k_{eff} prompt <1) tout se passe comme si la vie d'une génération était de quelques dizaines de secondes. Dans ces conditions, si la perturbation envisagée ci-dessus se produit, au bout d'une seconde la puissance n'a augmenté que d'un millième.

Des dispositifs commandent la chute de barres analogues aux barres de contrôle — mais ayant une capacité de capture beaucoup plus importante — dès que k_{eff} atteint une valeur considérée comme dangereuse, mais dans tous les cas inférieure à 750 « p. c. m ». Après la chute de ces barres, la chaîne est désamorcée.

Nous pouvons maintenant concevoir un premier schéma d'un système surcritique. Il doit comprendre :

— un bouclier biologique protégeant l'extérieur des radiations nucléaires émises;

— un réflecteur réduisant la proportion des neutrons qui fuient à l'extérieur;

— une ou plusieurs barres de contrôle permettant de faire varier k_{eff};

— un dispositif permettant d'extraire en permanence l'énergie thermique libérée par les réactions de fission.

Rappelons-nous que pour les phénomènes qui nous intéressent

— les noyaux composant un système peuvent être classés en 3 groupes :
les fissiles U 233 U 235 Pu 239
les fertiles Th 232 U 238
ceux qui ne sont ni fissiles ni fertiles,

— les principaux facteurs caractérisant le système sont :

- **k_{eff}**
- **le « taux de combustion »**
- **le facteur de conversion**
- **la taille du système**
- **l'enrichissement du combustible.**

comment fonctionne une pile atomique ?

Il faut maintenant, pour pouvoir poursuivre cette étude, prendre en considération les caractéristiques numériques des matières fissiles.

uranium naturel, ou uranium enrichi ?

Examinons d'abord quelques cas particuliers.

les réacteurs à uranium 238

Les neutrons de fission, qui ont une énergie moyenne de 2 MeV, peuvent provoquer la fission d'uranium 238, qui exige au moins 1,1 MeV. Mais il s'agit d'un processus rare : sa section efficace est de 0,2 barn, alors que celle du choc inélastique est de 2,5 barns; de plus, quand un neutron subit ce choc il perd plus d'un MeV, il tombe donc dans une zone d'énergie où il ne peut plus provoquer de fission. Il est évident que dans ce cas on ne peut pas avoir de réactions entretenues. Des considérations analogues sont valables pour le thorium 232.

les réacteurs à uranium naturel

Le même raisonnement ne peut pas être transposé sans modification au cas de l'uranium naturel, car ce mélange contient 7 p. 1 000 de l'isotope 235, qui est toujours fissile sous quelque énergie que ce soit. Les neutrons ayant subi le choc inélastique, bien qu'ayant une énergie inférieure à 1 MeV, peuvent provoquer des fissions avec les noyaux d'uranium 235. Cette possibilité est malheureusement limitée. Reprenons, en effet, la formule donnant le nombre de fissions par seconde et par centimètre cube :

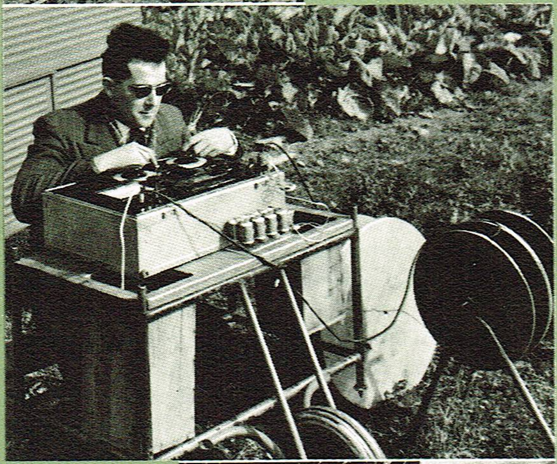
$$\varphi M \sigma_f.$$

Le nombre N de noyaux de matière fissile (c'est-à-dire de U 235) par cm³ est faible; d'autre part, aux énergies un peu inférieures au MeV, la section efficace de fission σ_f est également faible. On conçoit donc que les neutrons qui se trouvent dans la zone 10 keV - 1 MeV (par exemple après avoir subi le choc inélastique) produiront peu de fissions. Mais suivons-les : ils vont heurter des noyaux d'uranium et à chaque choc ils perdront un peu d'énergie; cette perte varie en raison inverse de la masse du noyau cible; elle est très faible quand ce noyau est lourd, par exemple l'uranium, elle serait très importante si les neutrons heurtaient des noyaux très légers, par exemple de l'hydrogène. Ils vont ainsi arriver dans

Prospection des minerais d'uranium.



avec le compteur Geiger

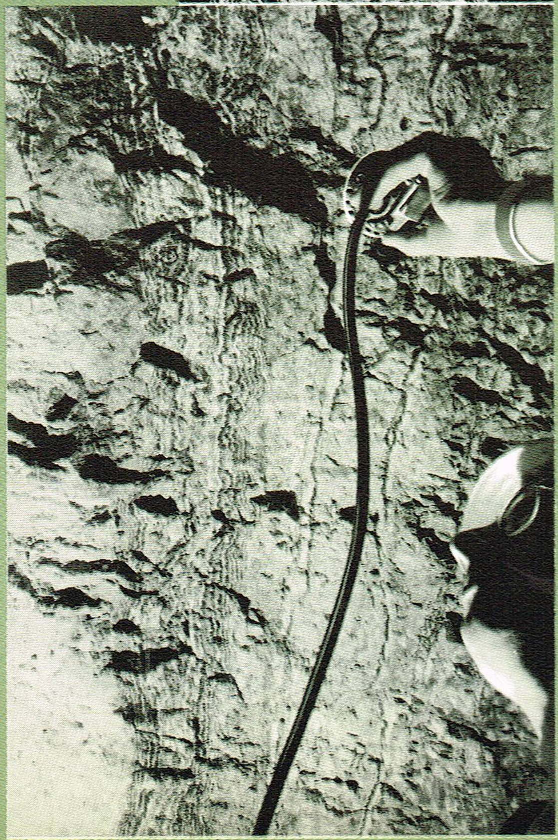


▲ par mesure de la résistivité du sol

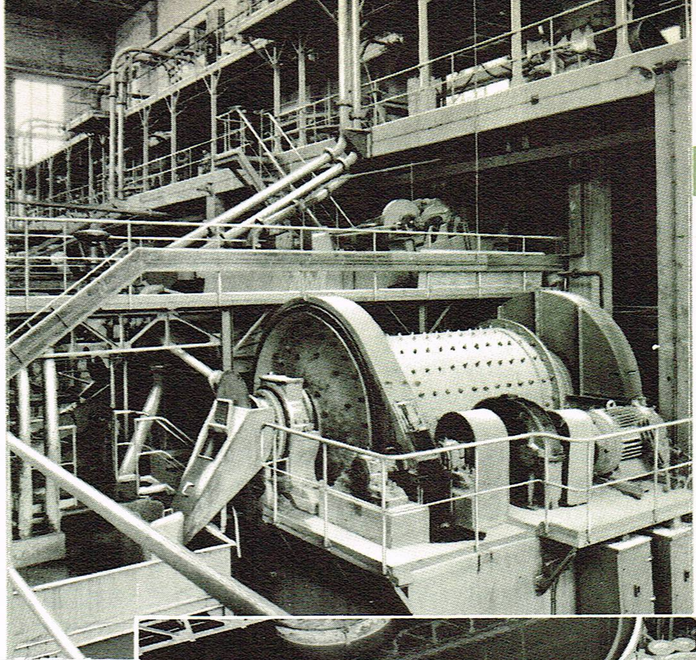


par dosage du radon du sol ►

filon uranifère ▼

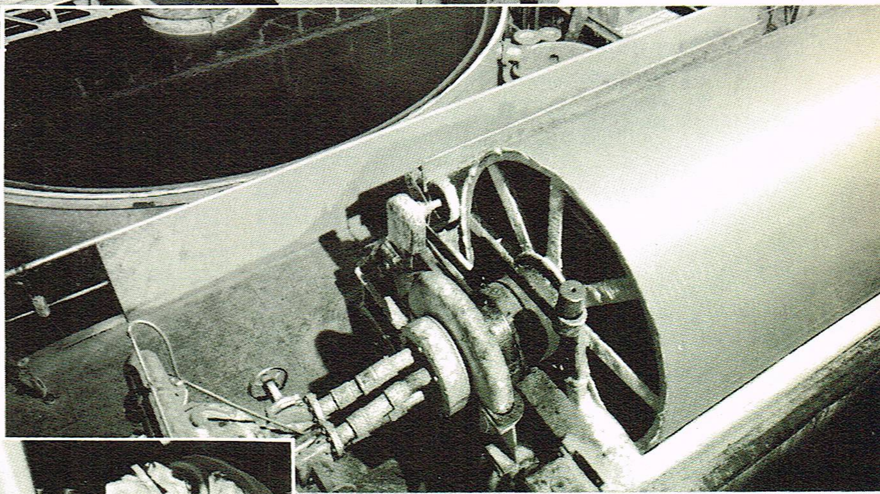


Traitement des minerais.



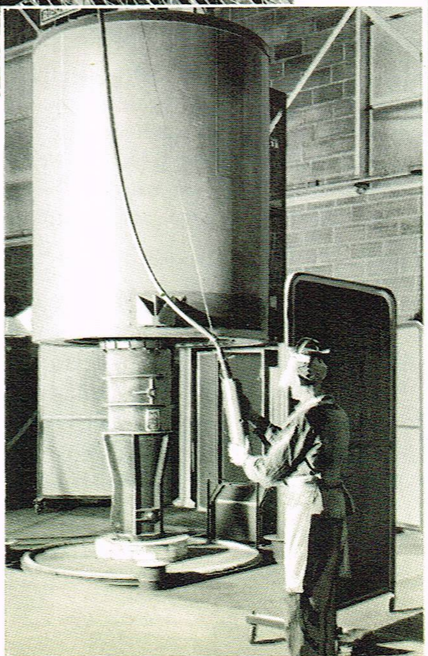
broyage

concentration chimique





◄ compactage du mélange minerai-magnésium dans le creuset



mise en place du four électrique autour du creuset



écroûtage d'un lingot d'uranium

Phot. Jahan (C. E. A.).

la zone 7 - 40 eV, qui présente des « raies de résonance », c'est-à-dire des intervalles d'énergie très faibles, à l'intérieur desquelles les sections efficaces sont très importantes. A chaque choc, le saut de vitesse sera très faible, et il y a finalement une probabilité importante pour qu'après un choc l'énergie du neutron soit à l'intérieur d'une raie de résonance; le neutron est alors presque sûrement absorbé avant d'atteindre le niveau thermique, il est perdu pour l'entretien de la réaction. Sans faire aucun calcul, il est donc à peu près évident **qu'il est impossible d'entretenir une réaction en chaîne dans un milieu contenant uniquement de l'uranium naturel.**

Pour sortir de cette impasse, il faut majorer N ou σ_f.

Pour majorer N, il faut augmenter la teneur du combustible en noyaux U 235 fissiles. On arrive à ce résultat en traitant un sel d'uranium naturel dans des installations extrêmement complexes, dites « usines de séparation isotopique » (une telle usine fonctionne depuis 1964 en France, à Pierrelatte). On a ainsi un uranium artificiel, dit « uranium enrichi », dont la teneur en noyaux fissiles peut atteindre 1 p. 100, 2 p. 100 ... 99 p. 100. **En fait, dès que la concentration atteint 10 p. 100, il est possible d'obtenir des réactions en chaîne.** De tels réacteurs sont dits « rapides », car la quasi-totalité des réactions nucléaires s'effectuent avec des neutrons d'une énergie supérieure à 100 keV.

neutrons thermiques, ou neutrons très énergiques

Pour majorer σ_f, il faut diminuer l'énergie des neutrons, les amener si possible dans la zone des énergies dites « thermiques », où la section de fission atteint 600 barns, en évitant qu'ils soient absorbés par les raies de résonance. Pour cela, il faut qu'au cours d'un choc leur variation d'énergie soit sensiblement supérieure à la largeur d'une raie de résonance. Cela exige que le choc ait lieu sur des noyaux légers. Dans le cas où la cible est un noyau d'hydrogène, certains neutrons peuvent passer de l'énergie de fission à l'énergie thermique en un seul choc; dans ce cas, ces neutrons évitent sûrement la capture par résonance.

De tels réacteurs, dans lesquels la majorité des fissions sont provoquées par des neutrons considérablement ralentis (leur énergie est une fraction d'électron-volt), sont dits « thermiques ». Ils devront comporter, en sus de la matière fissile, un corps étranger ayant des noyaux de faible masse, dont l'unique but est de ralentir les neutrons. Un tel corps s'appelle un « modérateur ».

pourquoi un modérateur ? comment le choisir ?

Une des premières propriétés de la matière constituant un modérateur est d'avoir une faible masse atomique que nous désignerons par A; de plus, il ne doit pas être très capturant pour les neutrons (σ_c faible); enfin, il faut que, par unité de volume, on ait le plus de chocs possible ($N\sigma_s$ élevé).

Les principaux modérateurs utilisés sont : l'eau légère, l'eau lourde et le graphite; le béryllium et son oxyde, la glucyne, conviendraient très bien, mais leur métallurgie n'est pas encore au point. Pour se ralentir jusqu'à l'énergie thermique — 1/40 eV —, un neutron subit quelques dizaines de chocs pendant un temps de quelques cent-millièmes de seconde.

La première conception qui vient à l'esprit est de mélanger intimement le modérateur et la matière fissile; on constitue ainsi des réacteurs dits « homogènes », mais il est évident que cette disposition ne conduit pas à l'efficacité maximale. Il y a toujours une partie des neutrons qui peut, même si le modérateur est excellent, tomber à l'intérieur des raies de résonance et être ainsi absorbée inutilement. Cette perte peut être réduite en délimitant à l'intérieur du réacteur deux groupes de zones : l'un contiendra uniquement du combustible, l'autre uniquement le modérateur. Avec cette disposition, les neutrons qui arrivent à un niveau d'énergie correspondant à une raie de résonance ne seront capturés que si, à cet endroit, ils se trouvent à proximité d'un noyau d'uranium. Ce phénomène ne sera évidemment qu'un cas très exceptionnel si la matière fissile se présente sous forme de blocs isolés au milieu du modérateur. De tels réacteurs sont dits hétérogènes.

Une fois le neutron suffisamment ralenti, il ne possède que l'énergie qui correspond à la température du modérateur. Il continue à se déplacer à énergie constante — en moyenne — jusqu'à ce qu'il disparaisse, soit en quittant le système, soit en étant absorbé par capture ou fission. Cette période qui ne dure que quelques millièmes de seconde s'appelle la « diffusion ».

Pour des raisons pratiques, les blocs de combustible ont presque toujours la forme d'un cylindre de révolution, c'est-à-dire de barreaux. Ils sont placés à l'intérieur de cavités de même forme, appelées « canaux », situées à l'intérieur du modérateur. Le diamètre de ces derniers est toujours sensiblement supérieur à celui des barreaux de combustible. Dans l'intervalle circule un fluide destiné à extraire l'énergie développée à l'intérieur des barreaux. Dans certains réacteurs, les blocs de combustible ont la forme de plaques.

Dans le cas où le modérateur est liquide, et où il peut servir de fluide réfrigérant, il n'est pas nécessaire de matérialiser les canaux, les barreaux de combustible sont simplement immergés dans le modérateur réfrigérant.

les gaines de protection

Afin de réduire le plus possible les risques biologiques dus aux rayonnements nucléaires, on enferme les barreaux de combustible dans une enveloppe métallique soigneusement obturée, dite « gaine ». Ce dispositif empêche les produits de fission de se répandre à l'intérieur du système dans le fluide réfrigérant. Le métal utilisé pour ce gainage devra satisfaire à certaines conditions : n'avoir aucune réaction chimique, ni avec le combustible nucléaire ni avec le réfrigérant, et surtout absorber très peu de neutrons; il devra également être bon conducteur de la chaleur, car l'énergie thermique créée au sein du

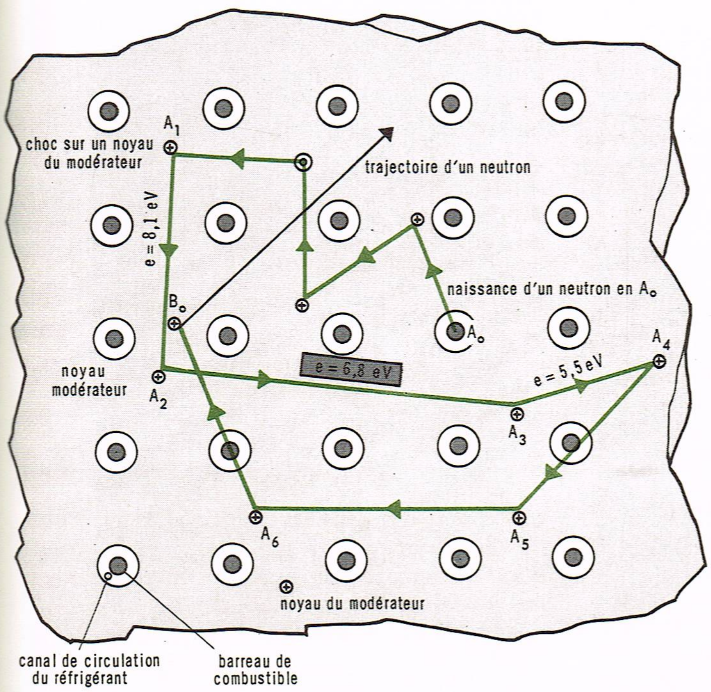
combustible devra pouvoir s'écouler dans le fluide réfrigérant en traversant le gainage.

Sous l'action prolongée du flux neutronique, les noyaux constituant certaines parties du système se déplacent les uns par rapport aux autres et arrivent à provoquer des déformations permanentes. Ce phénomène est particulièrement sensible sur les barreaux de combustible, car il est indispensable que les déformations restent en deçà de certaines limites, afin de ne pas gêner le déplacement de ces éléments à l'intérieur des canaux. Il faut également que ces déformations ne provoquent pas des « ruptures de gaines », sans cela certains produits de fission gazeux arriveraient à se répandre dans tout le système et à le contaminer. Ces déformations limitent le temps de séjour du combustible dans le système, donc le « taux de combustion ».

L'énergie thermique libérée dans un élément de volume est proportionnelle à

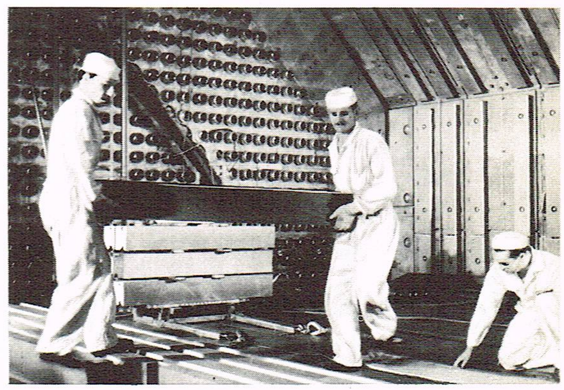
$$\varphi N \sigma_f.$$

Il y a intérêt à ce que dans chaque élément cette valeur soit égale au taux de combustion maximal admissible; pour qu'il en soit ainsi, il faudrait que le flux Φ soit constant dans tout le volume du système. Or, il n'en est pas du tout ainsi malgré les artifices utilisés pour se rapprocher de cette performance, ce qui s'appelle « aplatir le flux ». Au bout d'un certain temps de fonctionnement, les éléments de combustible qui sont au centre auront atteint par exemple les 2/3 du taux maximal, alors que les éléments latéraux n'en seront qu'au tiers. On a prévu des machines de déchargement et rechargement, qui servent non seulement à mettre en place la charge initiale, mais aussi à changer de place les barreaux de combustible, de manière que chacun atteigne le taux maximal. Ces machines sont très complexes et très encombrantes, elles doivent être aussi extrêmement robustes, car dans certaines installations elles sont appelées à fonctionner d'une manière quasi continue.



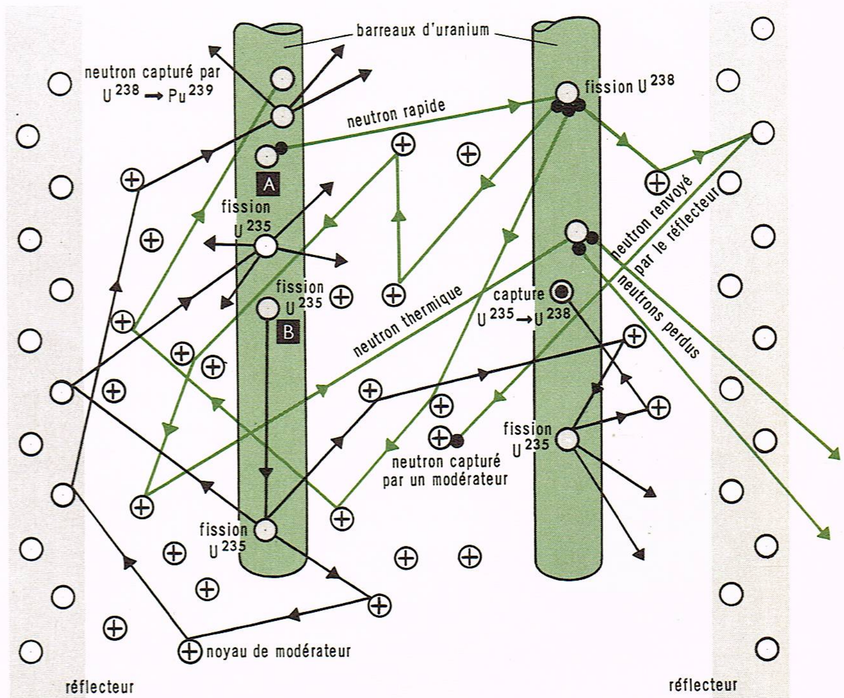
Coupe d'un réacteur uranium naturel - gaz - graphite par un plan perpendiculaire aux canaux.

Depuis A_0, où il a pris naissance, jusqu'à B_0, où il arrive à la vitesse qui correspond à l'équilibre thermique du modérateur, le neutron se déplace en zigzag. Cette phase, dont la durée est de l'ordre du dix-millième de seconde, s'appelle le « ralentissement ». Dans le graphite, la distance A_0B_0, comptée en ligne droite, est d'environ 50 cm.
Sur le parcours A_2 A_3, le neutron n'est pas capturé, bien que son énergie corresponde à celle d'une raie de résonance, parce qu'il ne rencontre pas de noyau d'uranium. Il en serait autrement si le modérateur et le combustible étaient intimement mélangés.
Après B_0, le neutron se déplace en zigzag, mais à énergie constante; il subit des chocs, puis disparaît, soit en quittant le système — fuite —, soit en étant capturé, ou en provoquant une fission.



Montage de l'empilement de graphite du réacteur de Marcoule.

Réactions neutroniques dans un réacteur thermique.



En un point A, un noyau d'U^{235} est fissé et produit un seul neutron, dont les descendants disparaissent rapidement. Le noyau fissé en B donne une descendance beaucoup plus nombreuse.

les réacteurs classiques

Nous voyons se dessiner deux grandes classes de systèmes : les thermiques et les rapides. Ces derniers utilisent obligatoirement un combustible enrichi et peuvent être surgénérateurs. Les thermiques seuls peuvent utiliser l'uranium naturel comme combustible.

Pour les modérateurs, on est limité à trois corps : eau naturelle, eau lourde et graphite. Notons que deux de ces modérateurs sont liquides; on pourra donc les utiliser comme réfrigérants. Quand on ralentit les neutrons avec le graphite, on peut refroidir les barreaux de combustible soit avec un gaz (gaz carbonique en général), soit avec un métal liquide (sodium, ou eutectique sodium-potassium). Ce dernier mode de refroidissement est indispensable avec les réacteurs rapides.

On peut également classer les réacteurs d'après l'utilisation prévue. Si l'on exclut les petits systèmes, qui ont toujours des puissances faibles inférieures au kilowatt et qui sont destinés à préciser certaines caractéristiques neutroniques d'un type de pile, on peut classer l'ensemble des systèmes divergents en deux groupes : les réacteurs de puissance et les réacteurs d'étude. Ces derniers sont principalement destinés à l'irradiation des matériaux; on cherche donc à obtenir un flux de neutrons important et les questions thermiques y jouent un rôle secondaire.

Par contre, les réacteurs de puissance sont destinés à produire de l'énergie mécanique, qui est en général transformée en électricité. L'énergie libérée par les fissions étant immédiatement dégradée en chaleur, la conversion sera limitée en rendement par le principe de Carnot; le réacteur nucléaire qui est l'équivalent de la source chaude devra fournir un fluide moteur à une température aussi élevée que possible. Nous allons voir que l'obtention de températures importantes se heurte à de graves difficultés, provenant principalement du fait que cette énergie thermique libérée à l'intérieur d'un corps solide ne peut se transmettre que par conduction; c'est ce phénomène qui limite la puissance.

les réacteurs à uranium naturel et graphite

Dans de tels réacteurs, la faible concentration des noyaux fissiles exige de réduire le plus possible les fuites et les captures parasites.

Deux possibilités pour diminuer l'importance des fuites : employer un réflecteur, ou augmenter les dimensions. Pratiquement, il faut utiliser les deux, et on arrive à des cœurs de l'ordre de 1 000 m³, c'est-à-dire à des cylindres ayant approximativement 10 m de diamètre et 10 m de hauteur. Le calcul montre que k_{eff} est maximal quand le rapport volumique du modérateur au combustible est de l'ordre de 60. Un tel cœur contiendra 200 à 300 t d'uranium métal et 1 500 t de graphite.

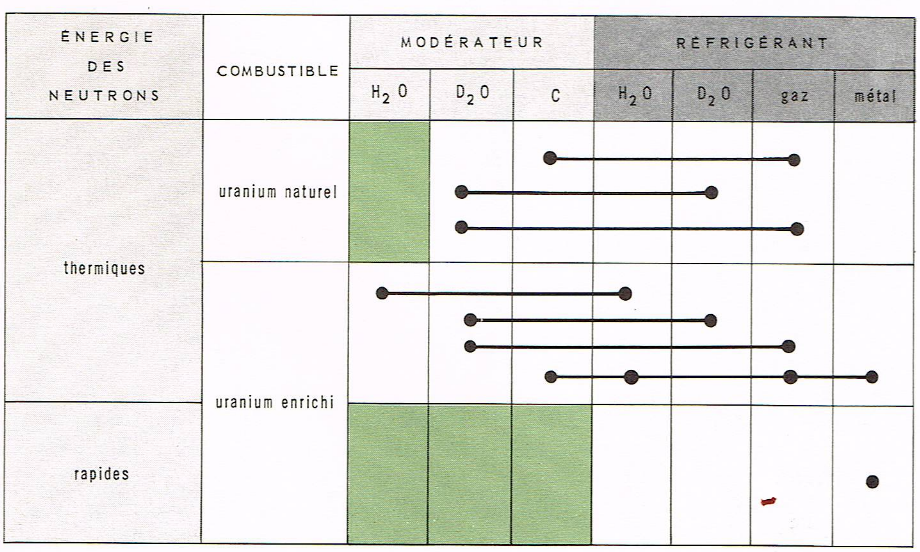
Pour réduire les captures parasites, il faut utiliser un modérateur très pur et n'admettre comme matériaux de structure que des corps très peu absorbants. Ces considérations excluent l'utilisation comme combustible d'un composé de l'uranium et nécessite des barreaux dont le diamètre ne soit pas trop faible.

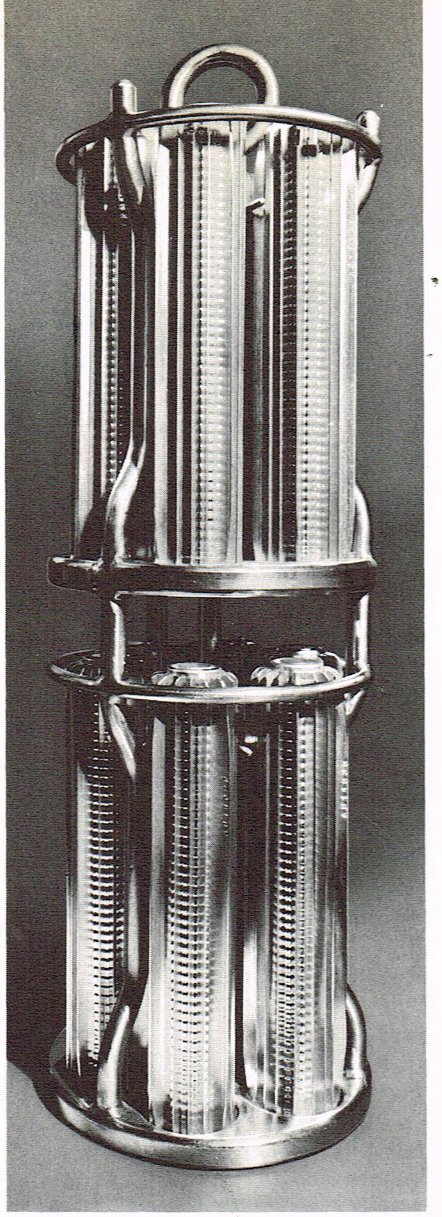
Pour le gainage des barreaux de combustible, on est pratiquement limité au magnésium : sa section de capture n'est pas très importante, et ses propriétés chimiques sont compatibles avec celles des autres corps avec lesquels il est en contact.

La puissance de tels réacteurs est aussi limitée par les caractéristiques métallurgiques de l'uranium et du magnésium, qui est le métal de gainage obligatoire. Pour ce dernier, on ne peut guère dépasser une température de 430 °C, ce qui fixe une température maximale pour le réfrigérant; celle de l'uranium ne doit pas dépasser 650 °C : au-delà, il se produit un gonflement qui provoque des ruptures de gaine. Ces deux températures conditionnent pratiquement le rayon des barreaux de combustible, qui est toujours de l'ordre de 10 à 12 mm. Des calculs simples montrent qu'on arrive ainsi à des puissances de l'ordre du mégawatt thermique par mètre cube de cœur.

Cette énergie thermique créée à l'intérieur du combustible devra être transférée au réfrigérant, qui dans ce type de réacteur est obligatoirement un gaz (pratiquement le gaz carbonique) : le coefficient de transfert est faible, et si on n'utilisait pas certains artifices, on aurait des installations ayant une faible puissance volumique, donc très onéreuses. Pratiquement, la gaine du barreau de combustible est garnie d'ailettes pour augmenter la surface d'échange, et le gaz est comprimé jusqu'à 12 kg/cm² (25 kg/cm² dans certaines centrales). Pour résister à une telle pression, l'ensemble du cœur est entouré d'un caisson métallique qui doit être constitué par des plaques ayant 10 cm d'épaisseur, ce qui pose des problèmes délicats de soudure. Les techniciens français ont mis au point une autre technique, qui se généralise au détriment de la précédente : le caisson est en tôles minces et les efforts sont encaissés par un corset de béton précontraint.

Le débit de gaz carbonique est de l'ordre de 5 t/s; il faut donc, pour le faire circuler, des ventilateurs très puissants, qui absorbent 3 à 4 p. 100 de la puissance thermique.

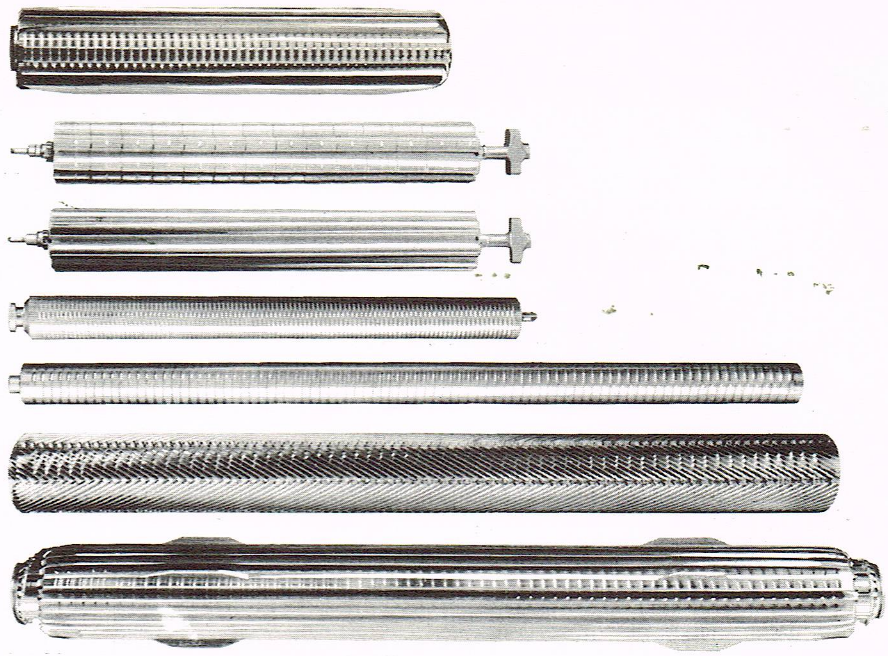




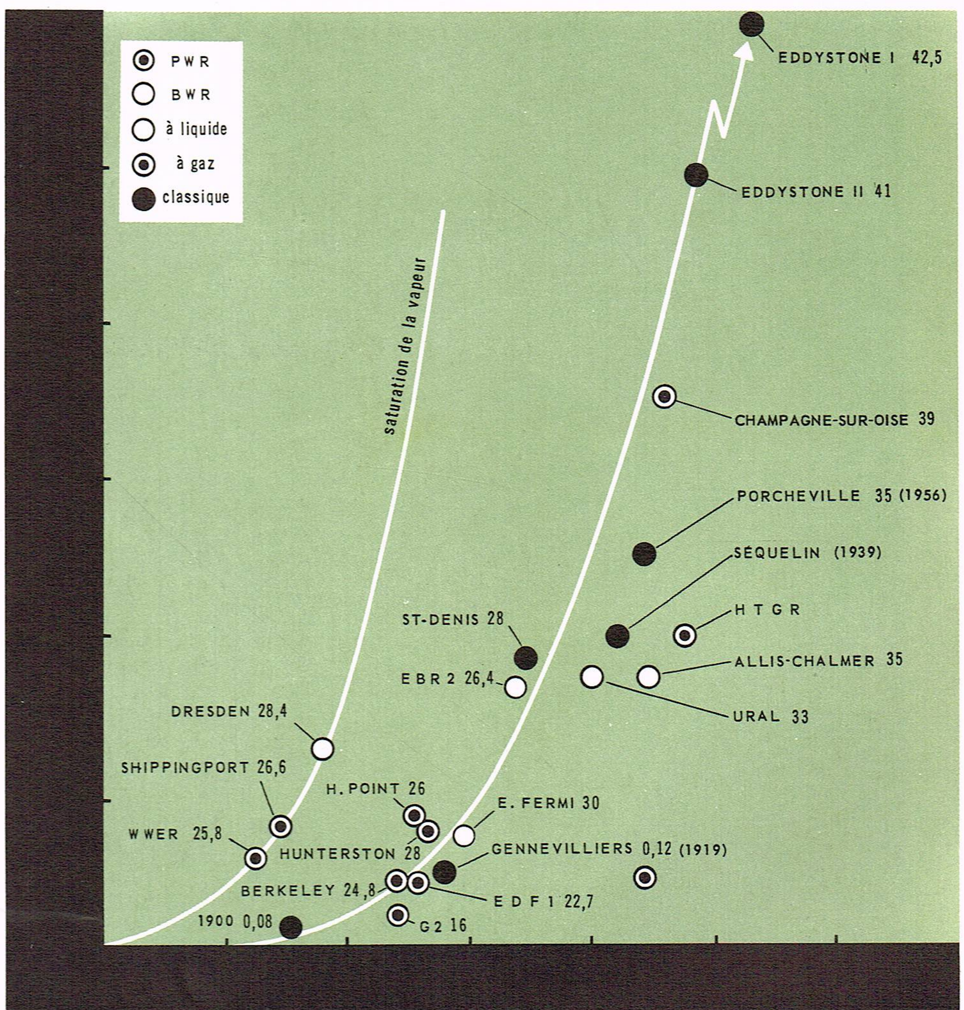
Eléments combustibles du réacteur G2 de Marcoule dans le bâti qui sert à les transporter.

Comme la température de la source chaude est limitée à 400 °C environ, il n'est pas possible d'utiliser des turbines à gaz, ce qui oblige à prévoir un échangeur de chaleur : organe compliqué et coûteux (la surface d'échange est de l'ordre de l'hectare) et qui réduit encore la température de 15 à 20 °C. La vapeur qu'il produit est donc peu surchauffée, ce qui exclut des rendements importants.

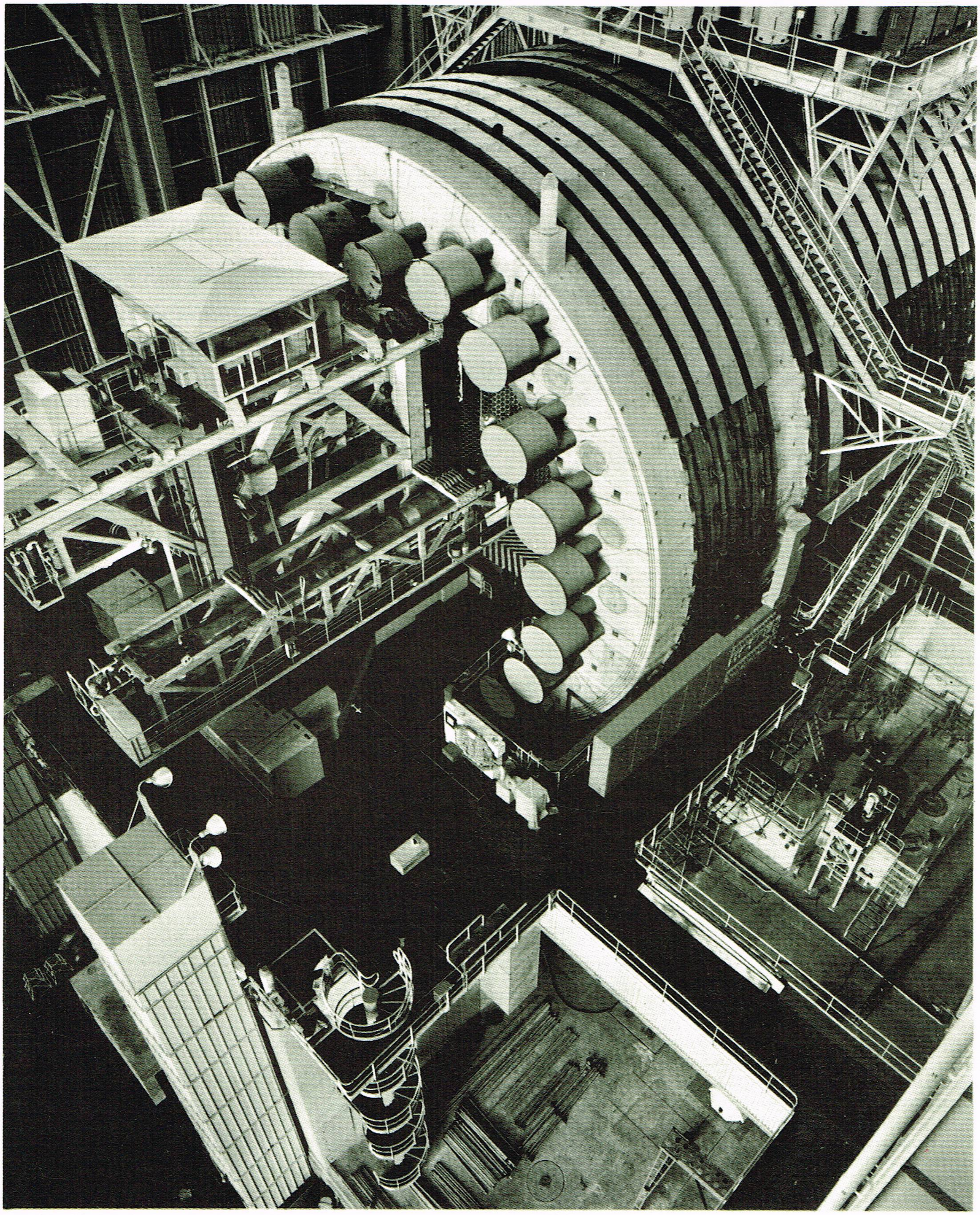
Dans les réacteurs utilisant l'uranium naturel, on arrive actuellement à des « taux de combustion » de l'ordre de 3 000 mégawatts-jours par tonne (MWj/t); mais les études en cours permettront certainement d'améliorer cette performance. Ces mêmes réacteurs ont un facteur de conversion de l'ordre de 80 à 85 p. 100 qui est sensiblement plus élevé que ceux qui sont obtenus avec les réacteurs des autres filières : c'est pour cette raison qu'ils sont souvent appelés *plutonigènes*.



La recherche du meilleur rendement, les impératifs d'ordre mécanique et thermique justifient ces formes apparemment fantaisistes données aux barreaux d'uranium qui constituent les éléments combustibles des réacteurs nucléaires. De haut en bas : barreau de la pile G2 (Marcoule); deux barreaux à profil dit « en cristal de neige »; les autres barreaux sont ceux des réacteurs « EL3 », « EL2 », « EDF2 » et « EDF1 ». Les chiffres indiquent le rendement de la centrale en pourcentage et l'année de mise en service.



Phot. Jahan (C. E. A.).



Un des réacteurs plutonigènes : G2, de Marcoule.

Phot. Jahan (C. E. A.).

les réacteurs à combustible enrichi

Nous venons de voir que presque toutes les caractéristiques d'une centrale à uranium naturel découlent logiquement de leur faible réactivité. Il n'y a pratiquement qu'un seul procédé pour augmenter cette réactivité : utiliser de l'uranium enrichi, ou encore un mélange d'uranium naturel et de plutonium.

Il suffit d'un enrichissement modeste (quelques pour cent) pour obtenir des accroissements très substantiels de ce paramètre. Le bilan neutronique devient alors beaucoup moins tendu, et il n'est plus indispensable de minimiser toutes les pertes. Il est alors possible de réduire les dimensions du cœur, de diminuer le rayon des barreaux de combustible, ce qui permet d'augmenter la puissance volumique, de choisir des matériaux de gainage (zirconium ou acier inoxydable) plus absorbants que le magnésium, mais acceptant des fluides réfrigérants ayant un meilleur coefficient de transfert que le gaz carbonique : l'eau, par exemple.

Des considérations très simples montrent que le nombre de combinaisons possibles pour organiser de tels réacteurs est, en fait, très limité.

Examinons d'abord le modérateur. Tant que la métallurgie du béryllium n'est pas au point, on n'en peut compter pratiquement que deux : l'eau naturelle et l'eau lourde, dont les avantages et les inconvénients sont résumés par les deux chiffres suivants : l'eau ordinaire est un ralentisseur quinze fois plus énergique que l'eau lourde, mais l'atome d'hydrogène est cent cinquante fois plus absorbant que celui de deutérium. **L'eau naturelle permettra donc de réaliser des cœurs très compacts, mais exigera un combustible plus enrichi que si l'eau lourde est choisie comme ralentisseur.**

Ces modérateurs liquides constituent d'excellents réfrigérants, et il n'est pas nécessaire de garnir d'ailettes les barreaux de combustible. Mais pour obtenir des rendements thermodynamiques intéressants, il faut augmenter le plus possible la température de cette eau; comme elle doit rester en phase liquide pour assurer efficacement son rôle de ralentisseur, il est nécessaire de réaliser une pression au moins égale à celle de la courbe de saturation; on arrive rapidement à une centaine de kilogrammes par centimètre carré. Alors intervient ici une limite technologique : cette eau sous pression devra être contenue dans des caissons d'acier dont les dimensions sont pratiquement limitées par des questions de facilité de transport et d'épaisseur de plaques que l'on sache souder. **Les réacteurs modérés à l'eau naturelle ont donné lieu à des développements très importants aux États-Unis dans deux filières un peu différentes : les pressurisés et les bouillants.**

La pression qu'il faut réaliser pour maintenir l'eau liquide conduit à des sujétions importantes; aussi a-t-on cherché des corps liquides riches en hydrogène qui aient des pressions saturantes très inférieures à celle de l'eau. Certains composés organiques correspondent à cette caractéristique et grâce à eux on a lancé une filière spéciale : les réacteurs organiques.

L'eau lourde est un moins bon ralentisseur que l'eau naturelle, mais comme la section de capture est pratiquement nulle, on peut concevoir des réacteurs à eau lourde qui divergent avec de l'uranium naturel. Ces réacteurs sont plus encombrants que ceux qui sont modérés à l'eau naturelle, et le prix de l'eau lourde est un sérieux handicap économique. Enfin n'oublions pas les réacteurs qui n'ont pas besoin de modérateurs : les rapides.

On utilise souvent l'uranium enrichi sous la forme de son oxyde UO_2, qui présente une excellente stabilité au contact de l'eau chaude, une très bonne tenue sous les rayonnements et une faible capture parasite; de plus, il ne fond qu'au-delà de 2 700 °. Il présente quand même quelques inconvénients : faible densité (10,8), très faible coefficient de conductibilité thermique. Enfin, comme il a une très mauvaise tenue mécanique, on doit l'utiliser sous forme de pastilles « frittées » enfilées dans un tube qui sert de gaine. Ce montage présente une « résistance thermique » importante à l'écoulement du flux de chaleur.

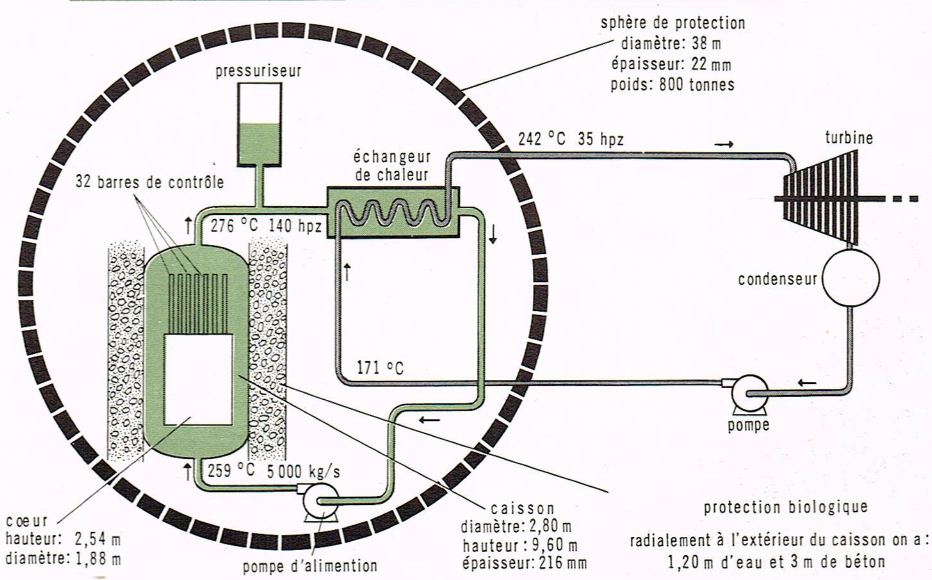
Le diamètre des barreaux de combustible dépasse rarement 1 cm; ils sont, en général, fixés par grappe à l'intérieur d'un parallélépipède métallique qui sert à canaliser le fluide réfrigérant. Avec toutes ces modifications, on compte obtenir des taux de combustion de 15 000 à 20 000 mégawatts-jours par tonne, et peut-être même beaucoup plus; quand le combustible est neuf, la réactivité à chaud est d'au moins 10 000 à 15 000 p. c. m., c'est-à-dire dix fois plus forte que celle des réacteurs à uranium naturel.

Nous allons terminer ce chapitre en donnant quelques indications sur les réalisations effectuées dans les différentes filières que nous venons de mentionner.

les « P.W.R. » (réacteurs à eau pressurisée)

Pour les raisons exposées ci-dessus, la vapeur produite sera donc toujours juste saturante sans aucune surchauffe, ce qui est néfaste pour les turbines; on essaie de

Schéma d'un réacteur pressurisé de type « Yankee » (400 MWth).



combustible : oxyde d'uranium enrichi, gainage en acier inoxydable, épaisseur : 0,5 mm, 76 assemblages contenant chacun 360 barreaux de 8 mm de diamètre

mettre au point des accessoires (cyclones) qui essorent l'eau contenue dans la vapeur aux étages intermédiaires de la turbine.

Pour tenir les pressions nécessaires, il faudra des caissons d'acier très épais. En fait, la taille actuelle des réacteurs modérés à l'eau naturelle est limitée par les possibilités de construction de ces caissons : de 1,5 à 4 m de diamètre et de 150 à 200 mm d'épaisseur. La hauteur est d'environ 10 m, car il faut tenir compte du débattement des barres de contrôle.

Pour éviter qu'elle ne devienne radioactive, l'eau doit être constamment purifiée; on ne peut dépasser 2 mg d'impuretés par litre. Pour des raisons économiques et de stabilité dans la régulation, on a intérêt a avoir une faible élévation de température du fluide réfrigérant dans le réacteur : de l'ordre de 20 °C; l'énergie développée est principalement utilisée pour la vaporisation. Compte tenu des caractéristiques physiques de l'eau, l'énergie de pompage est relativement faible.

Pour des raisons de sécurité, on enferme l'ensemble du circuit primaire — caisson, plus échangeur — dans une enceinte étanche, capable de supporter la pression provoquée par la rupture d'une partie de ce circuit, soit environt 3 kg par cm².

Ce type de réacteur a connu jusqu'ici un certain succès, car il équipe exclusivement tous les sous-marins et navires de surface à propulsion nucléaire; son faible encombrement et le bas prix de son modérateur en font une installation relativement économique.

les « B. W. R. » (réacteurs à eau bouillante)

Pour les réacteurs à eau pressurisée, la nécessité d'empêcher l'eau de se vaporiser exige d'une part un organe très important (le pressuriseur), et d'autre part entraîne la nécessité d'enfermer le tout dans un caisson capable de résister à des pressions très importantes.

On envisagea ensuite la réalisation de réacteurs où l'eau entrerait en ébullition à la partie supérieure du cœur; on est obligé d'augmenter un peu le volume pour loger la vapeur, mais on supprime le pressuriseur et la cuve supporte une pression déterminée par la seule température de l'eau. De plus, on peut supprimer l'échangeur de chaleur puisque l'expérience a prouvé que les produits de fission s'accumulaient dans la phase liquide. On gagne aussi sur les échanges thermiques, car le coefficient d'échange augmente sensiblement au début de l'ébullition. Mais il décroît brusquement quand une masse importante du liquide entre en ébullition; il ne faut pas dépasser plus de 1 à 2 p. 100 de vapeur en poids. Dans tous les cas, la puissance volumique est à peine la moitié de celles des « P. W. R. ». Le contrôle et la régulation sont plus délicats que ceux des « P. W. R. ».

Les centrales de ces deux types « P. W. R. » et « B. W. R. » diffèrent profondément des centrales à gaz-graphite : d'une part, les dimensions sont beaucoup moins importantes — la puissance volumique est cinquante fois plus importante avec les réacteurs à eau naturelle —; d'autre part, à puissance thermique égale, le poids de combustible est beaucoup plus faible ; enfin, l'utilisation de l'uranium enrichi fait passer au second plan les sujétions neutroniques. En fait, on dessine l'installation, puis on calcule l'enrichissement nécessaire pour obtenir le taux de combustion désiré.

Les deux types de réacteurs à eau naturelle sont considérés par un grand nombre de personnes comme ayant d'égales possibilités à l'avenir. Nous ne partageons pas entièrement cet avis et pensons que l'extension du programme de propulsion navale nucléaire a donné aux « P. W. R. » une avance que pourront difficilement rattraper les réacteurs bouillants.

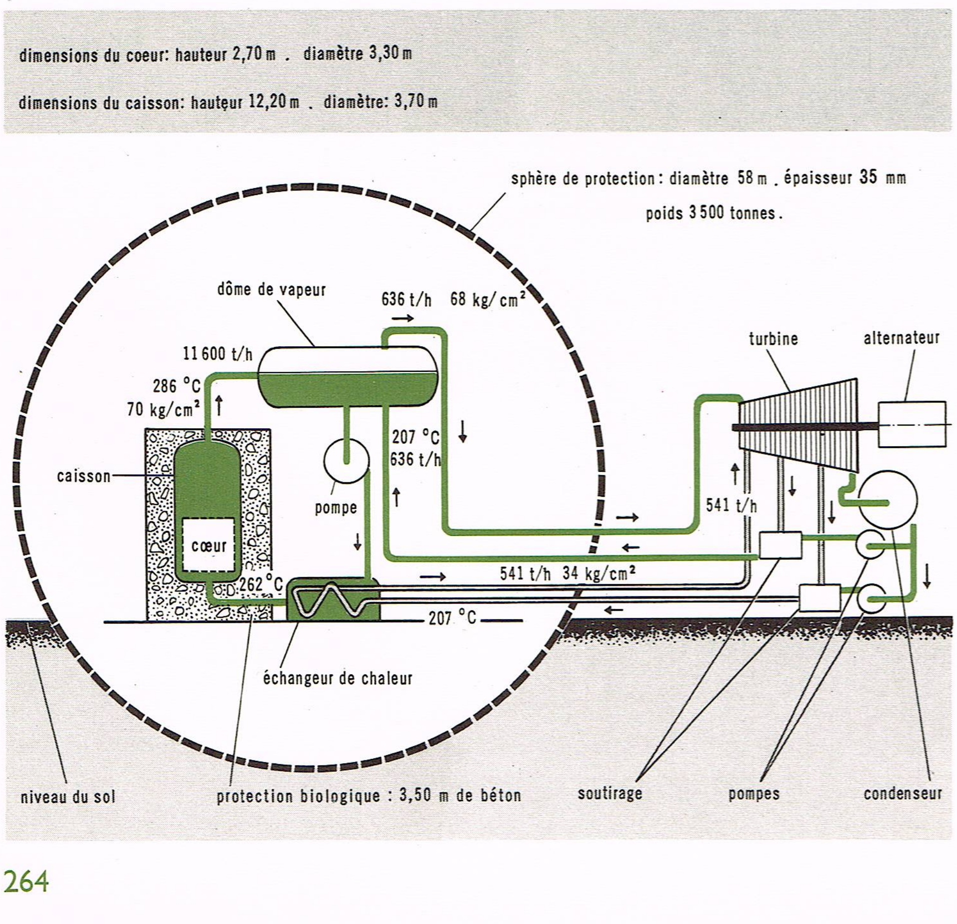


Schéma de la centrale de Dresden « B. W. R. »

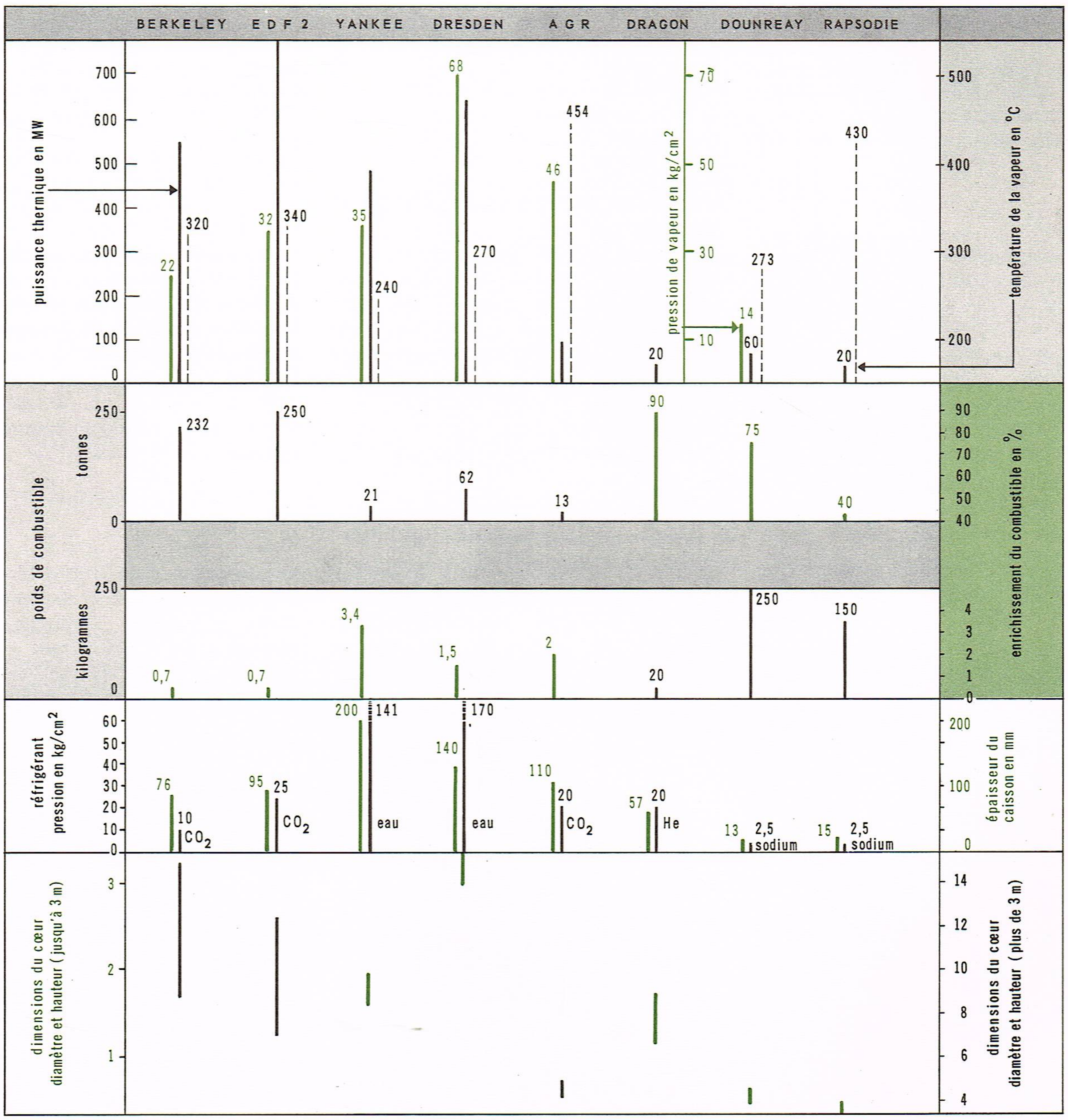
La sphère de protection a 58 m de diamètre, son centre se trouve à 17 m du sol, son épaisseur est de 35 mm.

les réacteurs de l'avenir

Nous venons d'examiner deux types de réacteurs; ceux qui sont modérés au graphite et ceux qui utilisent l'eau légère. Bien qu'aucun de ces réacteurs ne soit parfaitement au point, et que des progrès substantiels soient encore à faire, on a tendance à les considérer comme classiques, et de nombreux techniciens se penchent sur les futurs réacteurs, ceux que nous appellerons les « réacteurs avancés », et que nous allons décrire rapidement.

les réacteurs gaz-graphite

Comme dans le cas des réacteurs à eau naturelle, on peut, en utilisant un combustible enrichi, réduire les dimensions du cœur, augmenter la puissance volumique, accepter des fuites importantes et surtout envisager des taux de combustion également importants : disons 10 000 MWj/t. Pour améliorer l'évacuation de la chaleur, on augmente la pression du gaz carbonique (40 ou 60 kg/cm²); et en augmentant la différence de température entre la gaine et le gaz (550 °C) on peut se dispenser des ailettes, la température superficielle de gaine peut être de l'ordre de 1 000 °C. Le gainage présente évidemment des difficultés : on utilise l'acier inoxydable; le béryllium, aux propriétés extrêmement intéressantes, serait utilisé si on arrivait à mettre au point sa technologie. On espère faire encore un saut en température, obtenir



du gaz à 800 °C pour pouvoir faire fonctionner une turbine à gaz. Mais une autre limite intervient à ces températures : le gaz carbonique réagit sur le graphite; il faut changer de réfrigérant (en principe, on choisit l'hélium).

les réacteurs modérés à l'eau lourde

L'eau lourde, qui est un moins bon ralentisseur que l'eau naturelle, a, par contre, une section de capture presque nulle, et on peut concevoir des réacteurs modérés à l'eau lourde utilisant l'uranium naturel. Un inconvénient qui surgit est d'ordre économique; l'eau lourde coûte 300 F le kilogramme; comme il faut près d'un kilogramme d'eau lourde par kilowatt-heure installé, on voit que le modérateur intervient d'une manière très importante dans les investissements (à titre de comparaison, le kWh installé dans une centrale conventionnelle correspond à un investissement de 500 à 600 F).

Le modérateur étant liquide, on pourra l'utiliser comme réfrigérant, mais, comme pour les centrales à eau naturelle, il faudra le maintenir sous pression. La construction des cuves devant tenir la pression posera des problèmes de soudure très difficiles, car, à égalité de puissance, les dimensions des réacteurs à eau lourde sont plus importantes que celles des réacteurs modérés à l'eau légère. On préfère, en général, la solution dite « tube de force » (tube qui contient le combustible et qui résiste à la pression du réfrigérant); le modérateur qui se trouve à l'extérieur des tubes n'est pas sous pression.

On peut choisir, bien entendu, un gaz comme réfrigérant.

Les calculs neutroniques mettent en évidence qu'au départ on a une marge de réactivité importante, 4 000 à 5 000 p. c. m., et qu'il est possible, en faisant circuler le combustible de manière que chaque barreau ait la même irradiation, d'arriver à des burn-up de 10 000 MWj/t.

Nous indiquons ci-après quelques caractéristiques de deux réacteurs de puissance, modérés à l'eau lourde : un canadien, « Candu », et un français, « EL4 ».

« Candu » : 700 MWt; 188 tonnes d'eau lourde, dont 47 dans le cœur; 47 tonnes d'UO_2; 17 watts par gramme; l'eau lourde pressurisée à 96 kg/cm² sert de réfrigérant, sa température de sortie est 295 °C.

« EL4 » : 225 MWt, 122 tonnes d'eau lourde, dont 41 dans le cœur; 21 tonnes d'UO_2, 11 watts par gramme; refroidie par du CO_2 à 60 kg/cm²; température de sortie, 500 °C.

les réacteurs organiques

On a pensé utiliser comme modérateur et réfrigérant des corps contenant des atomes d'hydrogène, mais ayant des tensions de vapeur très inférieures à celle de l'eau; certains dérivés du phénol : diphényle, monophényle..., remplissent cette condition; tandis qu'au voisinage de la température critique de l'eau (375 °C), sa pression saturante est de 225 kg/cm², alors qu'elle ne dépasse pas 1 kg/cm² à la même température pour certains composés (santowax R). On imagine sans peine les simplifications alors procurées, d'autant plus que ces corps organiques n'ont pas d'affinité pour la plupart des matériaux utilisés, y compris l'uranium. Mais, malheureusement, ces corps se décomposent sous l'action des rayonnements et de la chaleur, ils peuvent s'enflammer vers 550 °C et — si des conditions très strictes de vitesse de circulation ne sont pas respectées — des dépôts se forment qui diminuent très sensiblement les coefficients de transfert, qui déjà sont toujours inférieurs à ceux de l'eau. Il est beaucoup trop tôt pour porter un jugement sur ce nouveau type de réacteur.

une solution d'avenir : les surgénérateurs

Nous avons vu comment certains réacteurs peuvent fabriquer des noyaux fissiles — en utilisant pour matière fertile de l'uranium 238 ou du thorium 232 — en nombre supérieur à celui de noyaux fissés détruits. On appelle de tels systèmes : surgénérateurs. La matière fertile y est disposée dans une couverture qui entoure le cœur.

A quelle condition un tel phénomène peut-il se produire? Classons les neutrons d'une génération d'une manière un peu différente de celle que nous avons adoptée jusqu'ici :

P_1 désignera le nombre de neutrons absorbés (capture + fissions) par le combustible;

P_2 désignera le nombre de neutrons capturés par les noyaux fertiles;

P_3 l'ensemble des neutrons qui ont fui, ou qui ont été capturés par des noyaux ni fissiles ni fertiles.

Le nombre de neutrons de la génération suivante sera égal à ηP_1, et si nous admettons que le réacteur est en régime permanent : $\eta P_1 = P_1 + P_2 + P_3$. Ou, en divisant par P_1 et en remarquant que la fraction P_2/P_1 est égale au facteur de conversion c :

$$\eta - 1 - c = P_3/P_1 > 0 \text{ entraîne } \eta > 1 + c.$$

Si l'on désire que, dans le réacteur, c soit supérieur à l'unité, il faut que $\eta \geqslant 2$. D'après les caractéristiques des noyaux, cette inégalité sera satisfaite au mieux quand on utilise un réacteur rapide alimenté par du plutonium. De tels systèmes sont dits surgénérateurs, ou « Breeder ».

Faut-il crier au miracle, en voyant la possibilité de fabriquer plus de combustible que ce que l'on consomme? Non. La philosophie de la question peut se résumer comme suit : l'homme sait aujourd'hui transformer en énergie utilisable une partie de l'énergie potentielle contenue dans un noyau « fissile », l'uranium 235. Les surgénérateurs transforment en noyaux fissiles deux noyaux qui ne l'étaient pas : l'U 238 et le Th 232. Ils facilitent l'utilisation d'un stock qui existe, mais ils ne créent pas d'énergie.

Pour des raisons économiques, la densité de puissance de ces réacteurs doit être très importante, et le refroidissement devra donc être extrêmement énergique : on est obligé d'utiliser des métaux liquides (sodium ou eutectique sodium-potassium) dont le coefficient de transfert est très supérieur à celui de l'eau.

Il est possible que ce type de réacteur soit appelé à un grand avenir; mais de nombreux problèmes sont encore à résoudre.

les réacteurs à surchauffe

En comparant les cycles thermiques des centrales nucléaires et ceux des centrales conventionnelles, on est immédiatement frappé par les différences de température (donc de rendement thermique) des vapeurs produites, surtout si on se limite aux réacteurs refroidis par de l'eau. Beaucoup de techniciens ont pensé surchauffer — par un dispositif nucléaire — la vapeur produite. En opérant ainsi, on augmente bien le rendement de l'installation, donc on diminue dans le prix du kilowatt-heure le terme coût du combustible, mais ce terme ne représentant qu'une faible partie du prix total, on gagne finalement très peu et il n'est pas du tout évident que ce gain ne sera pas absorbé par le coût des investissements supplémentaires.

des centrales bien spéciales : les explosions souterraines

Enfin, pour terminer ce chapitre, signalons qu'un grand nombre des difficultés signalées pour la réalisation des centrales nucléaires sont résolues par les explosions souterraines : avec des petites bombes à fission on creuse une grande caverne à 5 000 m de profondeur, on la remplit aux trois quarts d'eau, puis on y fait éclater une bombe H de 20 mégatonnes : toute l'eau est vaporisée, et on a ainsi une chaudière bien calorifugée; la vapeur remontée par les canalisations à la surface entraîne des turbines qui sont au niveau du sol, pendant plusieurs années. D'importantes sociétés ont pris des brevets concernant la réalisation de telles installations.

les centrales nucléaires

De telles installations diffèrent d'une centrale conventionnelle par l'organe qui produit la chaleur : dans un cas, c'est une chaudière classique; dans l'autre, un réacteur nucléaire.

quelques curieux problèmes technologiques

Nous avons vu que divers paramètres technologiques freinent dans les réacteurs la transmission de la chaleur : les centrales nucléaires disposent d'une source chaude à faible température, d'un rendement peu élevé, et d'une vapeur pas ou peu surchauffée, ce qui est nuisible à la bonne tenue des turbines. On essaie de pallier ces inconvénients, mais les moyens mis en œuvre jusqu'ici sont peu efficaces.

Une autre différence très importante provient des caractéristiques des produits de combustion : alors que dans les centrales conventionnelles ils sont rejetés dans l'atmosphère, il ne peut évidemment en être de même pour les produits de fission, presque tous radio-actifs. Beaucoup de précautions, toutes très onéreuses, doivent être prises pour que ces produits ne se répandent pas à l'extérieur. Il faut non seulement gainer le combustible, mais encore enfermer le plus souvent l'ensemble à l'intérieur d'un hémisphère dit « enveloppe de sécurité ». Même dans le cas des réacteurs de faible taille « P. W. R. » ou « B. W. R. », ces sphères ont des poids et des dimensions imposants : 50 m de diamètre, 1 000 t. Compte tenu des difficultés inhérentes à la réalisation de ces enveloppes, on y renonce souvent pour les réacteurs de grande taille, comme ceux

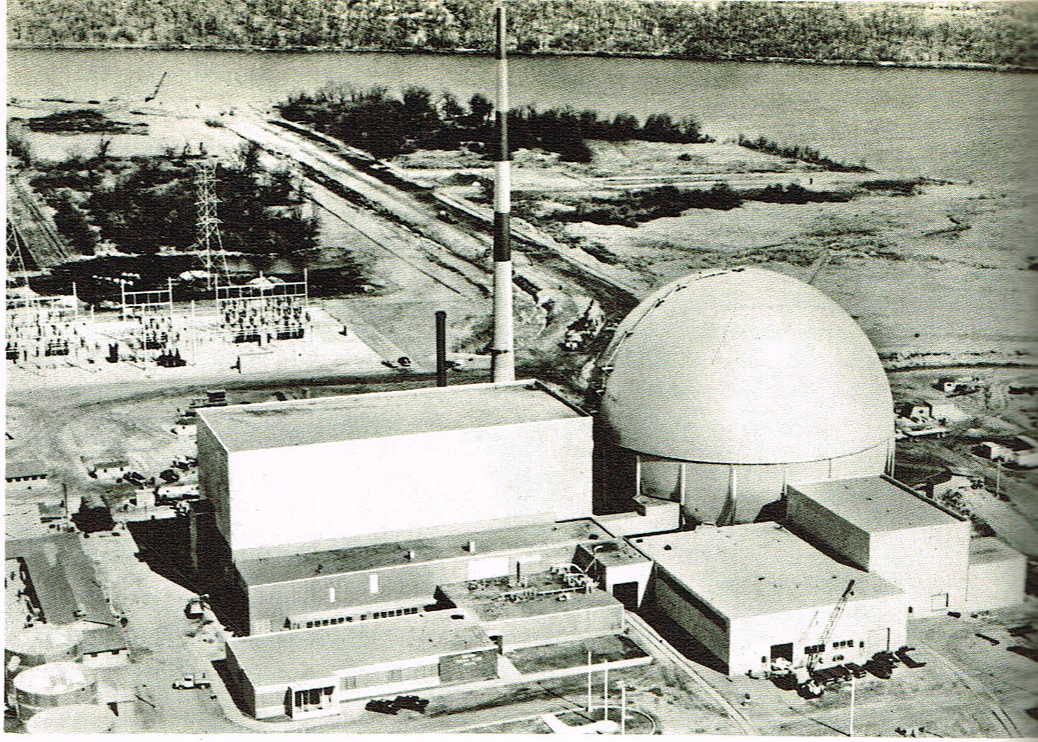
Centrale nucléaire de Chinon.



Phot. E. D. F.

de la filière uranium naturel gaz-graphite; il est alors nécessaire de soigner particulièrement l'étanchéité des canalisations de jonction entre le caisson et l'échangeur.

Même si ces produits de fission sont efficacement emprisonnés, il ne faut pas oublier qu'une partie des rayonnements qu'ils émettent : neutrons et rayons gamma, peuvent traverser d'importantes épaisseurs de matière et provoquer des ravages à l'extérieur de l'installation; ces rayonnements subsistent, atténués, après l'arrêt du réacteur. Afin de s'en protéger, l'ensemble du réacteur est entouré d'un « bouclier biologique » qui absorbe ces radiations : les caractéristiques varient d'un réacteur à l'autre, mais 20 cm d'acier et 2 à 3 m de béton sont des chiffres normaux.



Centrale nucléaire de Dresden.

et si la centrale explosait !

Malgré toutes les précautions prises, on peut se demander si la défaillance de certains appareils de contrôle ou encore des circonstances extérieures ne pourraient pas provoquer l'explosion d'un réacteur. La question a été souvent controversée, mais nous pensons qu'il faut y répondre par la négative : un réacteur ne peut pas exploser, au moins si on sous-entend par cette expression l'ensemble des phénomènes qui caractérisent le fonctionnement d'une bombe à fission. Ce qui ne signifie pas que certains incidents ne puissent pas avoir des conséquences presque aussi graves. Les effets d'une bombe tiennent à des causes différentes (élévation de température, pression de l'onde de choc), mais quels que soient les dégâts matériels provoqués par ces effets, ils sont relativement minimes comparés aux dégâts génétiques que peuvent provoquer les produits de fission. Ces derniers sont d'autant plus graves que les réacteurs se trouvent souvent à proximité de lieux présentant une forte densité de population. Mais il ne faut pas exagérer; certes, les installations nucléaires, comme beaucoup d'autres, présentent des risques certains, mais il suffit de le savoir et de mettre au point des dispositifs efficaces pour s'en protéger. *Dans ces conditions, on peut affirmer que les centrales nucléaires sont moins dangereuses que les conventionnelles, car elles ne diffusent pas en permanence dans l'atmosphère des produits aussi toxiques que l'oxyde carbonique ou sulfurique.*

Nous commençons à concevoir que les centrales nucléaires possèdent beaucoup d'organes qui n'ont pas d'équivalent dans les centrales conventionnelles. Si nous nous rappelons qu'il faut souvent changer de place le combustible à l'intérieur du réacteur, que les servomécanismes de régulation sont en général délicats, qu'il faut apporter un soin tout particulier à certains organes qui fonctionnent dans les zones irradiées du réacteur, et qu'enfin certains matériaux doivent être très purs, on comprend sans peine que les investissements nécessaires à l'érection d'une centrale nucléaire soient — à égalité de puissance électrique — beaucoup plus importants que ceux que nécessitent les centrales conventionnelles.

les centrales nucléaires sont-elles rentables ?

Nous avons vu que l'on ne peut pas optimiser simultanément les principaux paramètres : puissance par unité de volume, rendement thermique, nombre de canaux... Il faut donc faire un choix. Dans un grand nombre d'installations industrielles, on cherche principalement à rendre minimal le prix du kilowatt débité sur le réseau de distribution : bien que d'autres critères soient concevables, c'est celui-ci que nous allons retenir.

Le prix P du kilowatt-heure peut être considéré comme la somme de deux quantités, l'une afférente à la valeur des investissements, l'autre à l'exploitation de la centrale. Cette dernière peut, à son tour, se décomposer en deux : les frais fixes (que nous pourrons négliger, car ils ne représentent qu'une faible fraction des frais d'exploitation) et les dépenses de combustible.

Considérons une centrale destinée à fonctionner N années et ayant coûté en investissements A par unité de puissance électrique. On admet généralement que l'amortissement des dépenses initiales se fait en versant pendant chacune des N années une même somme, que nous poserons égale à Aj (j, taux d'amortissement). Si, au cours d'une année de fonctionnement, la centrale a fonctionné pendant f heures à la puissance nominale, la partie du coût du kilowatt-heure dépendant de l'amortissement sera Aj/f.

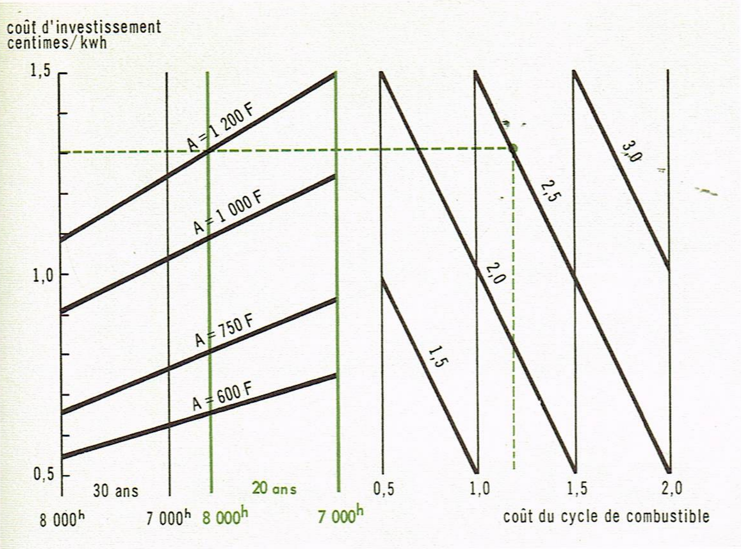
En désignant par p le coût du kilogramme de combustible, par B le « burn-up » (taux de combustion) et par ρ le rendement de l'installation, le combustible interviendra dans le prix du kWh pour $p/B\rho$, finalement :

$$P = \frac{Aj}{f} + \frac{p}{B\rho}.$$

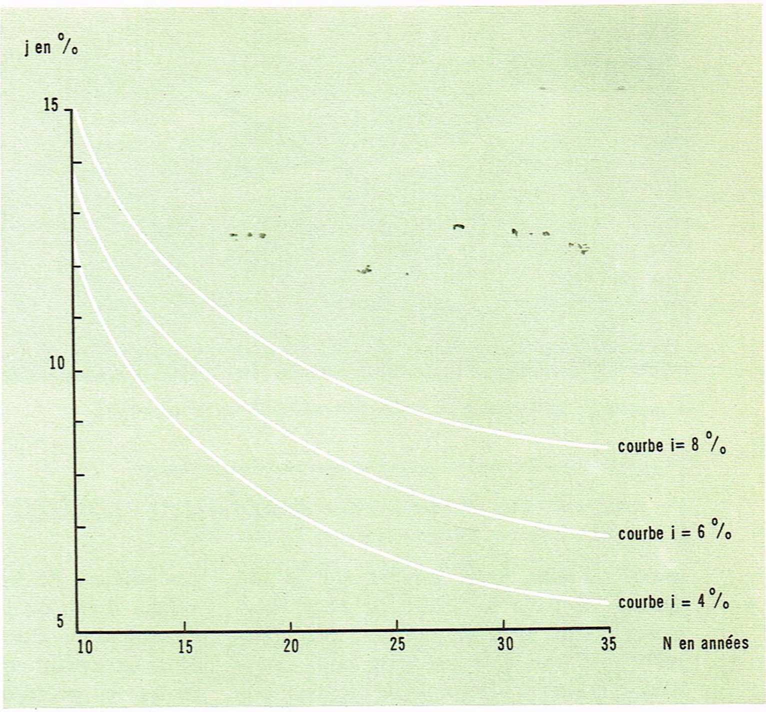
Nous allons examiner l'influence des différents facteurs.

D'abord ceux qui sont associés à l'investissement. Le paramètre A prix du kWh installé joue un rôle primordial, il varie sensiblement avec le type de la centrale

Phot. U. S. I. S.



Coût des investissements centimes/kWh. En fonction du coût A du kW installé, de la durée de vie, 20 ou 30 ans, et du temps annuel du fonctionnement, 7 000 ou 8 000 heures, on lit sur l'échelle de gauche la part du coût du kilowatt-heure dépendant de l'investissement. Par exemple : A = 1 200, 8 000 heures, 20 ans : 1,3 cent./kWh. A l'intersection de l'horizontale du chiffre et de la verticale du coût du cycle de combustible, on lit sur les lignes obliques le coût total du kilowatt-heure. Ex. : coût du cycle, 1,2; coût total, 2,5.



Variation du taux d'amortissement *j* en fonction du taux d'intérêt *i* et de la vie de l'installation N.

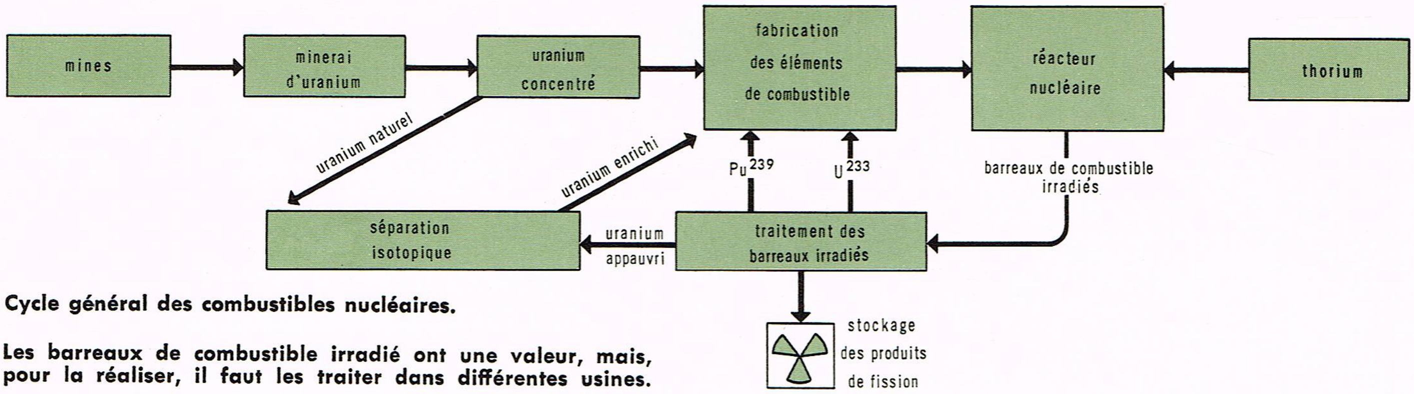
(uranium naturel ou enrichi) et avec sa puissance. Pour les centrales à uranium naturel que l'on construisait il y a quatre ans, sa valeur était de l'ordre de 2 000 F; aujourd'hui, pour Dungeness B, c'est un peu plus de 1 000 F; pour les projets qui verront le jour dans trois ou quatre ans, on pense arriver à 750 F. Pour les centrales à uranium enrichi, ce facteur varie de 750 à 600 F quand la puissance électrique passe de 400 à 800 MW; ce dernier coût est très voisin de celui des centrales au charbon. Le taux d'amortissement *j* dépend du taux annuel de l'intérêt *i* et de la durée de la vie de la centrale. Compte tenu de l'expérience acquise à ce jour, on ne doute plus que la longévité d'une centrale nucléaire soit du même ordre — trente ans — que celle des centrales classiques, car tous les incidents de fonctionnement qui se sont produits dans tous les pays sont mineurs et n'ont jamais mis en cause les principes de réalisation. En admettant qu'une centrale fonctionne 7 000 heures par an pendant 30 ans, et en prenant 6 % pour le taux annuel d'intérêt, on voit que la part du coût du kWh dépendant de l'amortissement passe de 1,04 centime à 0,62 quand le prix du kW installé varie de 1 000 à 600 F. Ces prix seraient majorés de 20 p. 100 si la vie de la centrale était réduite à 20 ans et diminués de 15 p. 100 si le temps de fonctionnement annuel était porté à 8 000 heures.

Voyons maintenant les facteurs attachés au fonctionnement. Le rendement y dépend principalement de deux facteurs : température de la vapeur et pertes par les fumées, qui varient en sens inverse l'un de l'autre quand on passe des centrales conventionnelles aux centrales classiques et finissent par se compenser.

Le taux de combustion est limité par deux conditions, l'une nucléaire, l'autre métallurgique. Avec l'uranium naturel, on peut obtenir des taux d'irradiation de 300 à 4 000 mégawatts-jours (80 000 kilowatts-heures environ) par tonne; avec un combustible enrichi, on peut envisager de 15 000 à 20 000 MWj par tonne (peut-être davantage, mais pour le moment on a peu d'expérience.

La valeur de *p* est la différence de deux termes, le prix du combustible neuf et celui du combustible irradié. La valeur du dernier est principalement représentée par les matières fissiles qu'il contient : plutonium et éventuellement uranium enrichi. Actuellement, un barreau d'U naturel gainé vaut environ 250 F le kilogramme et, après une irradiation de 3 500 MWj/t, il représente encore 50 F par kilogramme. Pour de l'U enrichi à 2 p. 100, ces chiffres sont respectivement 1 700 et 500 F/kg après une irradiation de 15 000 MWj/t. Mais il ne faut pas sous-estimer le caractère artificiel de certains de ces prix; les prix des matières fissiles sont fixés unilatéralement par les gouvernements en tenant compte de beaucoup de facteurs qui ne sont pas de nature économique. Pratiquement, pour les réacteurs type USA en cours de construction, on arrive, pour la part du prix du kWh dépendant du combustible — cycle du combustible —, à 0,9 centime; dans trois ou quatre ans on sera probablement à 0,7 centime. Pour les réacteurs à U naturel les plus évolués, le coût est aujourd'hui de 0,7 centime, et passera probablement à 0,5 dans quelques années. Nous avons cité beaucoup de chiffres, mais nous n'avons pas répondu clairement à la question : l'énergie nucléaire sera-t-elle compétitive?

Il y a trois ou quatre ans, personne n'avait répondu affirmativement, mais l'année 1965 a marqué aux Etats-Unis le début du « boom » nucléaire qui s'est concrétisé par des commandes extrêmement importantes de centrales nucléaires, et aujourd'hui, bien qu'aucune des centrales en fonctionnement ne soit compétitive, personne ne doute plus que dans peu de temps le cap de la compétitivité sera atteint et dépassé.



Cycle général des combustibles nucléaires.

Les barreaux de combustible irradié ont une valeur, mais, pour la réaliser, il faut les traiter dans différentes usines.

quelques industries amont et aval

Nous voulons enfin insister sur le fait qu'une centrale nucléaire ne peut fonctionner sans qu'il ait été créé préalablement une puissante industrie pour fabriquer et traiter les éléments de combustible.

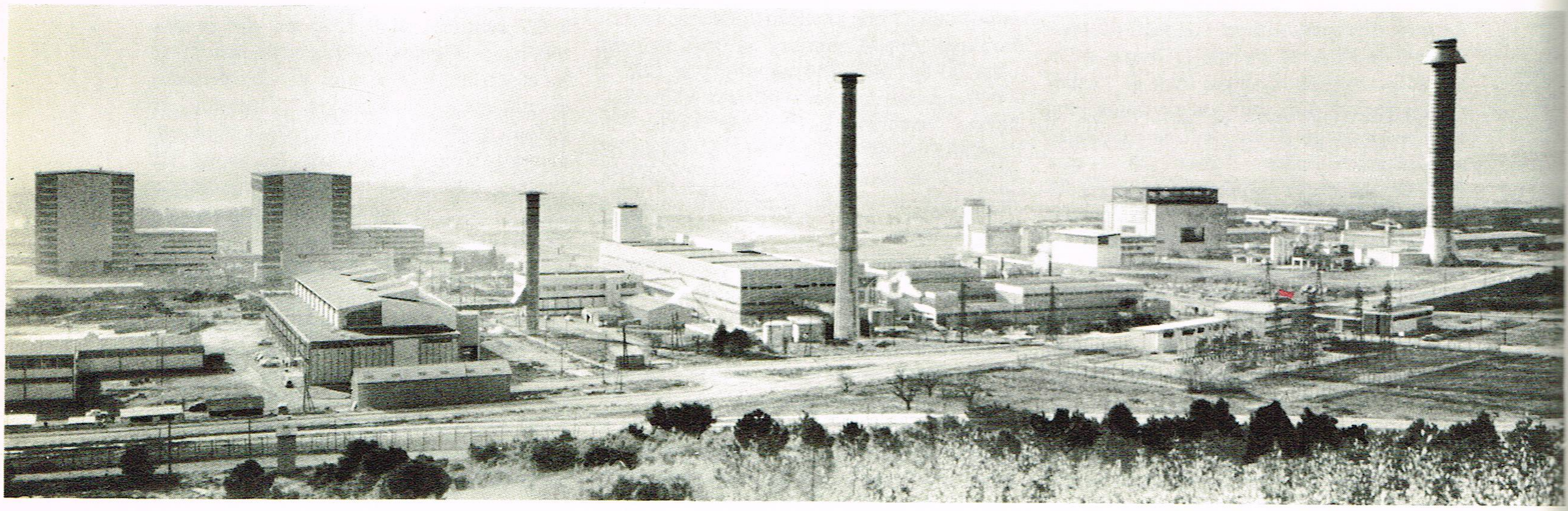
Le combustible, en effet, ne va pas directement de la mine à la centrale : quand le minerai est extrait, il faut le purifier, sa teneur ne dépassant que rarement 1 p. 100, puis extraire l'uranium-métal qui sera ensuite usiné et gainé.

Si l'on désire de l'uranium enrichi, le problème devient beaucoup plus compliqué. L'enrichissement s'obtient en faisant traverser des « barrières poreuses » à un sel d'uranium gazeux (pour rester dans une zone de températures basses on a choisi l'hexafluorure : UF_6). Les variations de concentration à la traversée d'une barrière sont très faibles; il faut prévoir des milliers de barrières; toute trace de vapeur d'eau est prohibée à l'intérieur de l'installation, dont le volume total est plus important que celui d'une cathédrale; les problèmes d'étanchéité et de lubrification des machines prennent ici une importance considérable... **finalement ces usines sont extrêmement onéreuses.**

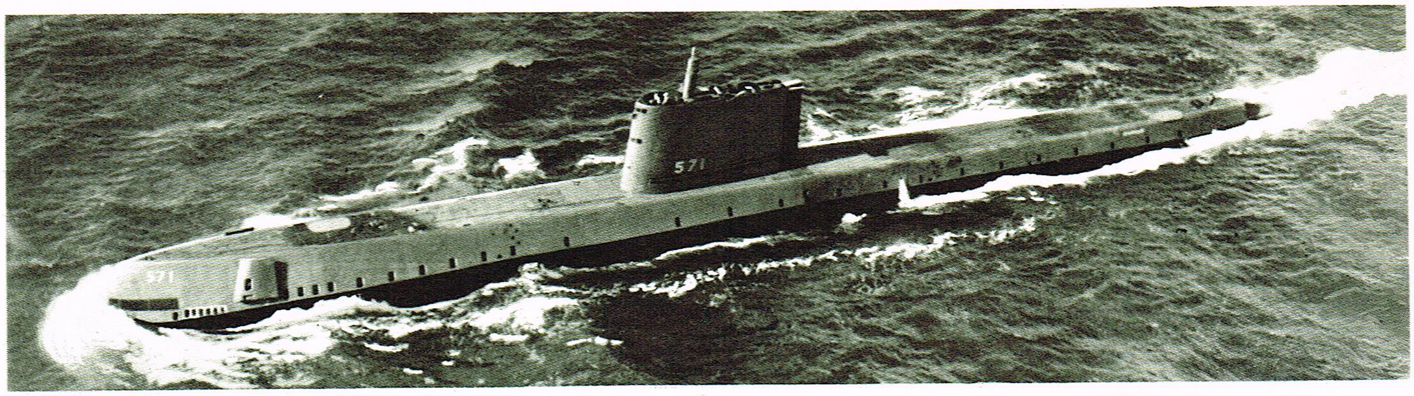
Mais il n'y a pas que le combustible qui exige une préparation spéciale : certains corps parasites qui ont une section de capture importante doivent être pratiquement exclus, leur concentration doit être de l'ordre d'une partie pour un million; cette pureté « nucléaire » majore le prix de certains matériaux : graphite, zirconium...

L'eau naturelle utilisée dans certains réacteurs doit être aussi très pure. La fabrication de l'eau lourde D_2O exige des installations très spéciales, dont la mise au point fut fort longue.

Il y a aussi les industries d'aval. Les barreaux — ou plaques — de combustible irradié ont une certaine valeur, mais elle ne devient tangible que quand on a pu extraire et séparer le plutonium, l'uranium 233, l'uranium appauvri, le thorium et les produits de fission. Toutes ces opérations exigent de vastes laboratoires de chimie, extrêmement compliqués par le fait que tout doit être commandé à distance pour éviter la contamination due aux produits radio-actifs. Mais dans tous les cas, il faudra stocker les résidus; la technique générale consiste à les transformer en composés solides, présentant le volume minimal, qui seront ensuite enfermés dans des containers étanches, placés dans des souterrains ou dans des mines désaffectées.



Centre de production de plutonium de Marcoule. *Phot. P. Jahan (C. E. A.).*



Sous-marin américain *Nautilus*, à propulsion atomique.

les navires à propulsion nucléaire

Un inconvénient inhérent aux réactions de fission est la nécessité de recourir à des blindages — bouclier biologique — épais pour protéger le personnel contre les radiations nocives émises. Si ces sujétions ont posé à l'avion atomique des problèmes qui n'ont pas été résolus, il n'en est pas de même dans la marine, où les impératifs de poids sont moins sévères.

Dès 1948, aux Etats-Unis, sous l'impulsion du dynamique amiral Rickover, des études furent lancées pour réaliser le *Nautilus*, premier sous-marin atomique. A cette époque, aucun réacteur n'avait produit de l'électricité, et les problèmes à résoudre étaient considérables.

D'une part, il a fallu mettre au point un réacteur très compact, pour parvenir à être logé à bord d'un sous-marin (on a choisi un type nouveau : « P. W. R. », qui a été mis au point par Westinghouse), et d'autre part, il a fallu étudier pour le gainage des éléments de combustible la métallurgie du zirconium.

Plusieurs prototypes à terre furent construits et essayés, avant d'arriver à une version quasi définitive. *Nautilus*, qui jauge 3 200 t (98 m de longueur et 8,5 de largeur), fit ses premiers essais en mer le 17 janvier 1955. Son autonomie, extrêmement importante, atteint presque 200 000 km. Sa puissance et sa vitesse sont tenues secrètes, mais celle-ci excède certainement 20 nœuds (37 km/h). On a peu de renseignements sur le coût de cette réalisation, qui est probablement de l'ordre de 500 millions de francs.

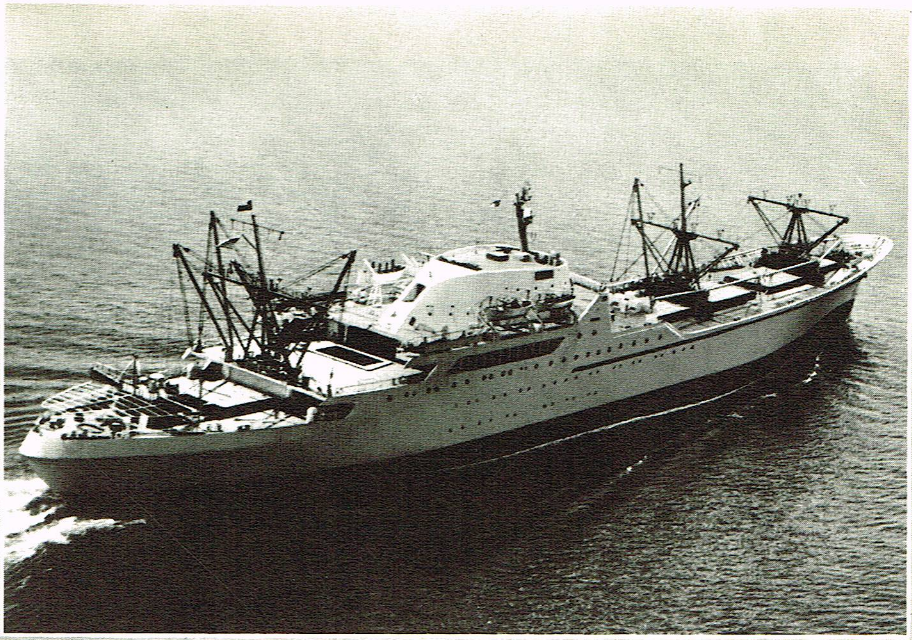
Parallèlement, mais avec quelque dix-huit mois de décalage, un sous-marin atomique d'un autre type, le *Seawolf*, fut réalisé aux Etats-Unis; il était équipé d'un réacteur à « neutrons intermédiaires », et refroidi par du sodium liquide. Les études et réalisations furent confiées à la General Electric Company. Lors des



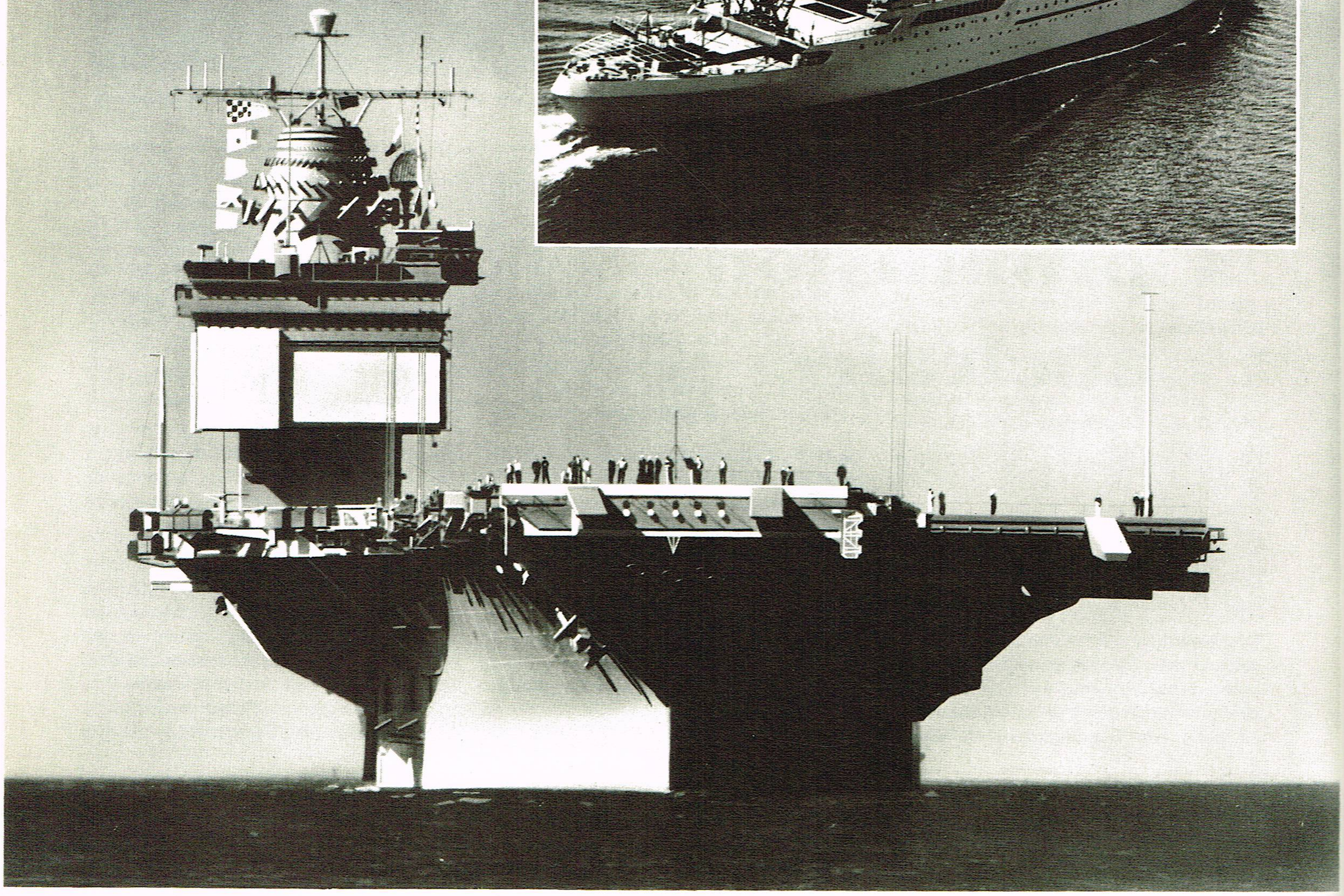
Phot. A. F. P., B. S. I.

Brise-glace soviétique *Lénine*, à propulsion atomique.

Cargo américain *Savannah*, à propulsion atomique.



Porte-avions américain *Enterprise*, à propulsion atomique.



essais, le réacteur donna entière satisfaction, mais des fuites se produisirent au surchauffeur de vapeur.

La marine américaine argua de cet incident, qui pourtant ne mettait pas en cause la valeur du réacteur, pour décider que tous les navires militaires nucléaires seraient équipés de « P. W. R. ».

A l'heure actuelle, près d'une centaine de sous-marins atomiques sont en service ou en construction.

Ils appartiennent d'après leurs missions à quatre ou cinq types différents; un des plus importants est le groupe du *George Washington* (5 600 t), équipé de seize fusées « Polaris » (portée 1 500 miles), qui peuvent être lancées le sous-marin étant en plongée.

Toujours sous l'impulsion de l'amiral Rickover, on essaya d'introduire la propulsion nucléaire dans les navires de surface. A ce jour, six bâtiments sont en chantier ou en service :

— un croiseur lourd de 35 000 t, le *Long Beach ;*

— deux porte-avions de 85 000 t;

— trois destroyers de 7 500 t.

Les avantages de ces navires sur leurs homologues ayant un moyen de propulsion classique n'ont pas été estimés suffisants pour faire oublier leurs prix. Il est donc probable que l'extension de la propulsion nucléaire aux navires de surface ne se fera qu'avec une sage prudence.

Phot. U. S. I. S.

qu'a-t-on fait dans les marines militaires des autres pays ?

En France, les études sur le sous-marin atomique remontent à 1954; elles ont porté d'abord sur un réacteur à uranium naturel (seul combustible disponible en France) qui aurait équipé un sous-marin expérimental : le « Q 244 ». Les travaux furent arrêtés en 1958, lorsque apparut la possibilité d'obtenir des Etats-Unis de l'uranium enrichi. Un accord militaire franco-américain assure à la France la matière première du combustible pour une installation prototype à terre réalisée à Cadarache.

La Grande-Bretagne a profité de ses accords avec les États-Unis pour construire deux sous-marins atomiques : le *Dreadnought* et le *Valiant*, lancés respectivement en octobre 1960 et en juillet 1966.

On a peu d'informations sur les réalisations soviétiques, mais l'U. R. S. S. possède sans doute plusieurs sous-marins atomiques.

Il est à peu près évident que les avantages conférés aux sous-marins par ce nouveau mode de propulsion sont tellement importants, tant aux points de vue autonomie, vitesse, que possibilité de rester très longtemps immergé, que ce type de navire se développera rapidement.

et la marine marchande ?

Le premier navire atomique civil fut le brise-glace russe *Lénine*, lancé le 5 décembre 1957. Il fit ses essais en septembre 1959. C'est un bâtiment de 16 000 t, équipé de trois réacteurs pressurisés; un est mis en réserve, les deux autres qui sont en service développent en tout 40 000 CV.

Le second navire fut un cargo mixte américain : le *Savannah*, de 22 000 t; il fut lancé en juillet 1960, et a commencé à naviguer en septembre 1962. Il peut transporter 10 000 t de fret à la vitesse de 21 nœuds. Son prix est très élevé : plus de 150 millions de francs (1963).

Le troisième, l'*Otto Hahn*, est un cargo allemand, mis en chantier en décembre 1962 : c'est un navire de 10 000 t, qui filera 17 nœuds avec un réacteur « organique » qui fournira 10 000 CV sur l'arbre.

D'autres projets ont vu le jour dans différents pays : à notre connaissance, aucun n'est en cours de réalisation.

En France, un concours a été ouvert en 1957 pour étudier un avant-projet de pétrolier atomique : trois types de réacteurs ont été étudiés : « P. W. R. », « B. W. R. » et « à gaz », mais aucune conclusion officielle ne fut tirée de ce concours.

En Grande-Bretagne, un comité fut mandaté par le gouvernement en mai 1959 pour examiner différents projets, présentés soit par les services officiels, soit par des groupements privés. Après une première élimination, ce comité retint trois systèmes : « P. W. R. », « B. W. R. » et organique. Une étude plus poussée montra que les deux derniers étaient mieux adaptés à la propulsion navale que le premier! Cette conclusion semble curieuse, car tous les navires qui existent sont équipés de pressurisés.

On pourra peut-être discuter encore longtemps pour justifier les mérites de telle ou telle filière de réacteurs. Nous pensons, contrairement à ce qui se passe pour les sous-marins, que la propulsion nucléaire n'apporte actuellement que de minimes avantages aux navires de commerce. On ne peut donc établir une comparaison qu'en faisant intervenir des caractéristiques économiques et nous estimons que notre expérience presque insignifiante en la matière rend tout calcul illusoire. De toutes façons, les navires marchands nucléaires sont très chers par rapport aux conventionnels.

Ce que nous pouvons connaître des avantages comparés des divers types de réacteurs est à notre avis très bien résumé dans une boutade de l'amiral Rickover : *Chaque nouvelle filière présentée comporte généralement l'ensemble des caractéristiques suivantes : « c'est simple ; — c'est petit ; — c'est bon marché ; — c'est léger ; — cela peut être construit très vite ; — il n'y a pas de dépenses de développement ». Par ailleurs, les « P. W. R. » sont caractérisés par les propriétés suivantes : « C'est en retard sur les délais ; — cela demande beaucoup de « développement » sur des parties en apparence très simples ; — c'est gros ; — c'est compliqué ; — c'est lourd. » Mais nous nous permettons d'ajouter : « Cela existe. »*

Devant le coût de navires marchands nucléaires, faut-il irrémédiablement condamner ce nouveau mode de propulsion? Non pas, mais il faut prendre conscience du fait qu'on ne le rendra pas compétitif à coups de petits perfectionnements, mais plutôt en modifiant profondément toute l'architecture navale. Nous ne citons pour mémoire que les cargos sous-marins, qui ont donné lieu à des discussions passionnées.

On n'arrivera probablement à rien d'intéressant si l'on remplace, dans une coque classique, un moteur Diesel par un réacteur nucléaire. Mais peut-être, avec un tel réacteur, sera-t-il possible de concevoir des cargos à très grande vitesse.

en cas d'incident ou d'accident

On ne peut terminer ce tour d'horizon sans parler des accidents éventuels, et des risques spéciaux que les navires nucléaires peuvent faire courir.

Dans le cas d'un naufrage en pleine mer, les produits de fission peuvent se répandre. On peut certes imaginer qu'une très grande dilution réduise considérablement leur nocivité; mais si des courants sous-marins les entraînent près des côtes? Beaucoup de personnes pensent qu'en prévoyant des structures « anti-collision », le réacteur ne sera pas endommagé et conservera les produits de fission dans les gaines des éléments de combustibles. C'est possible, mais qu'arrivera-t-il quand la corrosion de l'eau de mer aura fait son œuvre? De plus, on ne peut pas *a priori* exclure un accident grave dans un port ou au voisinage de côtes ayant une forte densité de population. L'accident du malheureux sous-marin américain *Thresher* n'a pas apporté des réponses claires à ces questions.

Jusqu'ici, on a essayé de résoudre ces problèmes par un système d'assurance spéciale. Les risques à garantir sont tels qu'il est presque indispensable que les gouvernements interviennent. Les Etats-Unis accordent une indemnité gouvernementale de 500 millions de dollars, en sus de ce qui est garanti par les compagnies

d'assurance, pour tout dommage qui serait provoqué par le *Savannah*. De plus, ce navire ne peut se rendre dans les eaux territoriales d'un Etat qu'avec l'autorisation du gouvernement de cet Etat. Il va de soi que les critères qui feront donner ou refuser l'autorisation seront plutôt fondés sur des considérations politiques ou économiques, que sur des études de sécurité. Mais enfin et surtout, le problème le plus délicat en cas de sinistre sera de définir les dommages, certaines lésions biologiques pouvant se révéler plusieurs semaines, si ce n'est plusieurs mois ou plusieurs années après le sinistre.

la propulsion nucléaire spatiale



L'idée d'appliquer l'énergie nucléaire à la propulsion des avions a été émise pour la première fois en 1930 par Robert Esnault-Pelterie, dans son livre *l'Astronautique*, où il écrit :

« La conclusion s'impose donc que l'asservissement de l'énergie intra-atomique nous rendrait possible et même facile la visite du système solaire tout entier, mais qu'eu égard aux abîmes qui nous séparent des autres systèmes stellaires, même les plus proches, l'exploration de ceux-ci sera éternellement interdite à l'homme. »

des moteurs nucléaires sur les avions

Il fallut attendre 1946 pour voir l'U. S. Air Force lancer les premiers programmes d'avions à propulsion nucléaire, et 1948 pour que l'A. E. C. (Atomic Energy Commission) s'intéresse à la question. En 1950, tous ces programmes furent fondus en un seul : l'A. N. P. (Aircraft Nuclear Propulsion) et son premier objectif fut de commander des réacteurs nucléaires destinés à équiper le bombardier « B. 36 ».

Ce programme subit beaucoup de vicissitudes : en 1953, les commandes furent résiliées; en 1954, on reprit l'affaire; en 1956, on déclara que ce n'était pas urgent; en 1959, le point de vue est à nouveau modifié, quand on apprend que l'U. R. S. S. lance des études; enfin, le 29 mars 1961, dans son message au Congrès, le président Kennedy décida, bien qu'on ait déjà dépensé plus d'un milliard de dollars, que le programme ANP serait arrêté et les crédits supprimés.

Mais ces dépenses ne furent pas totalement inutiles : les résultats obtenus par les constructeurs débordent largement la propulsion des avions et peuvent s'étendre à celle de l'engin balistique ou des véhicules spatiaux, et à toutes les applications où il est nécessaire d'avoir un réacteur léger et de faible encombrement. On peut affirmer qu'au cours de ces douze ou treize années d'étude, les Etats-Unis ont édifié la technologie des réacteurs à haute température; progressivement, les températures de sortie du réacteur sont passées de 250 °C à 900 °C, et des progrès énormes ont été faits dans l'étude des matériaux résistant aux hautes températures.

Indiquons rapidement dans quelles voies ces études furent conduites. Au début, on a simplement placé un réacteur nucléaire en « sandwich » entre le compresseur et la turbine d'un turboréacteur d'avion, c'est donc l'air préalablement comprimé qui servait de réfrigérant au réacteur. On réalise ainsi une machine qui fonctionne en circuit ouvert; cela s'appelle aussi un « cycle direct ». Les études furent exécutées par la General Electric Company, qui fabriqua successivement les réacteurs « HTRE-I », « HTRE-II » et « HTRE-III », qui fonctionnaient avec des neutrons lents.

Le principal objectif était d'arriver à des températures du même ordre que celles des turboréacteurs d'avion : 1 000 °C

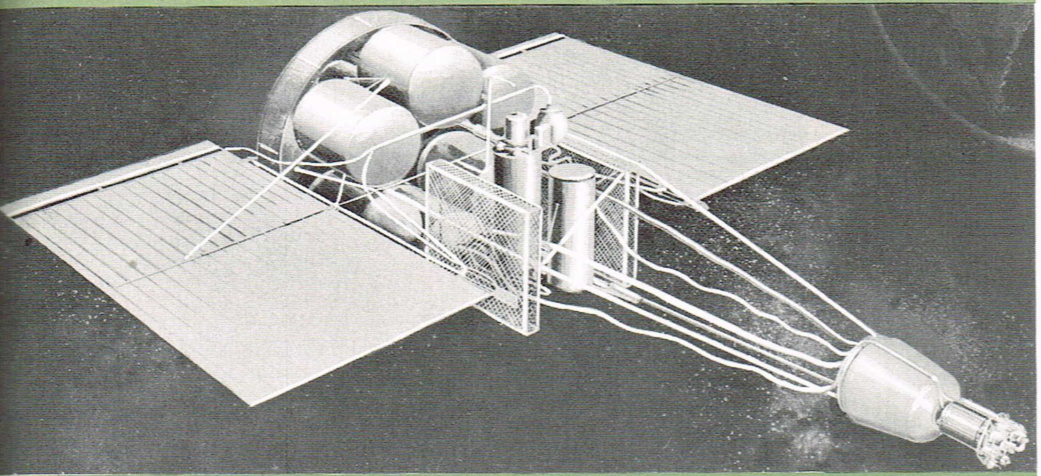
environ. Malgré des avantages certains, on ne peut songer à utiliser dans un cycle direct des réacteurs rapides, car l'air se prête mal aux échanges thermiques. On peut également concevoir un avion équipé d'un réacteur nucléaire central, qui fournirait de la chaleur à plusieurs turboréacteurs; le transport de chaleur devrait presque obligatoirement s'effectuer avec un métal liquide : cela s'appelle un « cycle indirect » (ou « fermé »), *a priori* beaucoup plus compliqué qu'un cycle ouvert.

Un problème très difficile à résoudre était celui de la protection des équipages; de nombreuses études furent faites pour essayer de réduire le poids du bouclier biologique, et le gain que permet de réaliser le cycle indirect fait que, dans certains cas, il peut être retenu. Mais il ne faut pas oublier les risques que ferait courir le réacteur nucléaire dans le cas d'un accident. **Finalement, la propulsion nucléaire n'a pas rencontré beaucoup de succès dans les applications aéronautiques.**

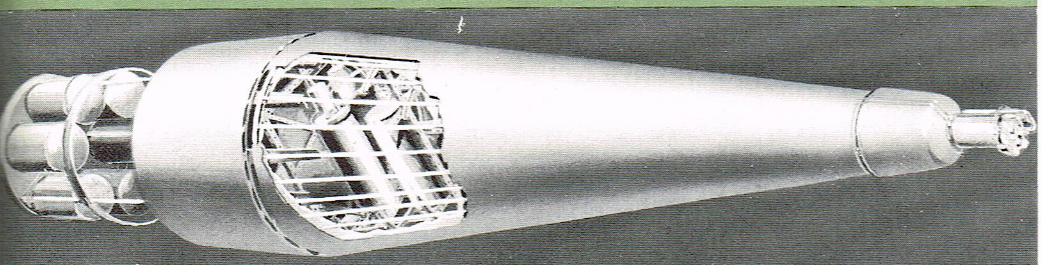
Si le but n'a pas été atteint, il faut surtout en accuser l'évolution trop rapide de la technique des « plus lourds que l'air »; l'aérodynamique et les « moteurs » ont fait des progrès tels que les avions ont été pratiquement rayés des programmes militaires et remplacés par des aérodynes sans voilure : missiles, engins ou fusées.

◀ **Maquette du nez d'une fusée équipée d'un générateur nucléaire « SNAP » et trois applications de ce genre de réacteur. Mais ce ne sont là que des extrapolations!**

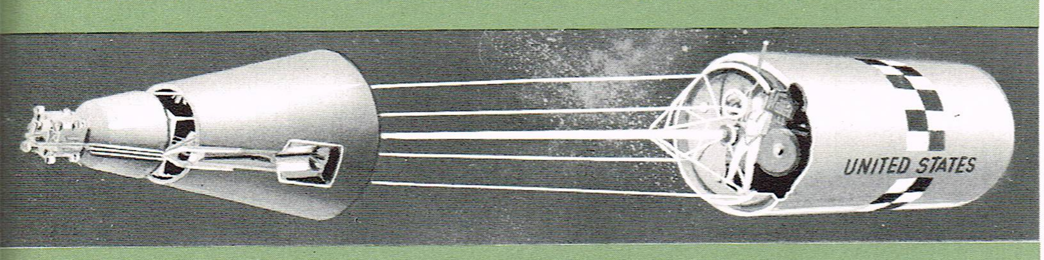
(D'après « Atomics International », U. S. A.)



Propulsion d'un engin interplanétaire par un « SNAP 8 ».



Le « SNAP », utilisé comme générateur d'énergie électrique pour la propulsion atomique.



Astronef propulsé par un « SNAP 2 ». L'engin est lancé par des propulseurs chimiques. Puis le bloc propulseur se sépare de l'habitacle, auquel il reste relié par des câbles et par un tube télescopique.

(D'après « Atomics International », U. S. A.)

missiles et fusées

Parmi les projets les plus importants qui se sont substitués au programme ANP, il faut signaler le projet « Pluto », qui correspond à un missile, et le projet « Rover », à une véritable fusée nucléaire.

L'objet du projet « Pluto » était l'étude et la réalisation d'un statoréacteur nucléaire destiné à propulser un missile dans les couches basses de l'atmosphère. La première réalisation fut le réacteur « Tory II A », qui n'était pas destiné à voler et qui, au mois de mai 1961, exécuta des essais au sol concluants : il a atteint 50 000 kW pendant deux cents secondes. A ce moment, l'ensemble du projet avait consommé 200 millions de dollars et l'AEC décida de passer à un modèle plus perfectionné, le « Tory II C », destiné à une présentation en vol. Des crédits de 300 à 400 millions de dollars étaient prévus pour ce second stade.

Malgré la quasi-certitude de réaliser les performances prévues pour « Tory II C » (engin pouvant porter une charge thermonucléaire et se déplacer à des vitesses de plusieurs « mach » avec une grande autonomie), malgré la difficulté de concevoir un système d'armes efficaces contre de tels engins se déplaçant à basse altitude, il semble que la réalisation de « Tory II C » ne soit pas considérée comme de première urgence, et peu de crédits lui sont maintenant consacrés.

La raison profonde de ce désintéressement, peut-être passager, est qu'actuellement le gouvernement américain donne la priorité aux programmes spatiaux. Les véhicules appelés à circuler dans les couches denses de l'atmosphère céderont peu à peu le pas à ceux qui peuvent se soustraire à l'attraction terrestre : aux fusées. La course au prestige semble se substituer à la course aux armements, mais les fusées peuvent également avoir des applications militaires.

Progressivement, en pensant aux performances déjà réalisées par les V2 allemands, les techniciens se sont orientés vers les fusées, véhicules qui transportent le car-

burant et le comburant nécessaires et qui se propulsent par réaction en éjectant les produits de la combustion. La valeur de la poussée est égale à mv, où m est le débit de gaz par unité de temps et v la vitesse d'éjection. Le calcul montre que quand une fusée a quitté la zone d'attraction terrestre, ses principales performances (entre autres son rayon d'action et sa vitesse maximale) sont proportionnelles à la vitesse d'éjection v, elle-même proportionnelle à $\sqrt{T/M}$ (T et M sont la température absolue et la masse moléculaire des gaz éjectés) : on cherche donc à augmenter T et à diminuer M.

Dans une fusée chimique, les hautes températures sont obtenues par la chaleur libérée par des réactions de combustion qui sont toutes des réactions de synthèse, la masse moléculaire des corps éjectés ne peut donc pas être très faible. On arrive à des vitesses de l'ordre de 3 000 m/sec., la température maximale que l'on peut atteindre étant limitée par les caractéristiques des matériaux.

Un des avantages de la fusée nucléaire est de pouvoir opérer avec des fluides ayant une masse moléculaire M beaucoup plus faible, car il n'est plus nécessaire de provoquer une réaction de synthèse pour avoir de la chaleur. On pourra, par exemple, avec un réacteur chauffer de l'hydrogène ($M = 2$) et atteindre des vitesses de l'ordre de 10 000 m/s; pratiquement, cette vitesse ne sera limitée que par les températures maximales que pourront supporter les éléments de combustible du réacteur.

Tout cela suppose que la fusée est libérée de l'attraction terrestre. Or, pour la libérer, il faudra lui appliquer une poussée supérieure à son poids, qui est lui-même supérieur à celui de l'appareil propulsif, c'est-à-dire à ωW, en désignant par W la puissance et par ω le poids spécifique du réacteur :

$$p > \omega W = \omega \frac{1}{2} m v^2 = \frac{\omega v}{2} m v = \frac{\omega v}{2} p,$$

ou encore $\omega < 2/v$.

Cette inégalité conduit, pour les vitesses d'éjection envisagées, à des performances très difficiles à obtenir au sol avec un réacteur nucléaire, mais facilement envisageables quand la fusée est éloignée de la Terre. On conçoit ainsi des véhicules à plusieurs fusées (plusieurs étages) mises en route successivement, les premières, à énergie chimique, étant utilisées pour sortir de la zone d'attraction terrestre. Les calculs complets montrent que l'utilisation de fusées nucléaires, pour les étages supérieurs, permet de réduire le nombre total d'étages et autorise donc la construction de véhicules moins complexes.

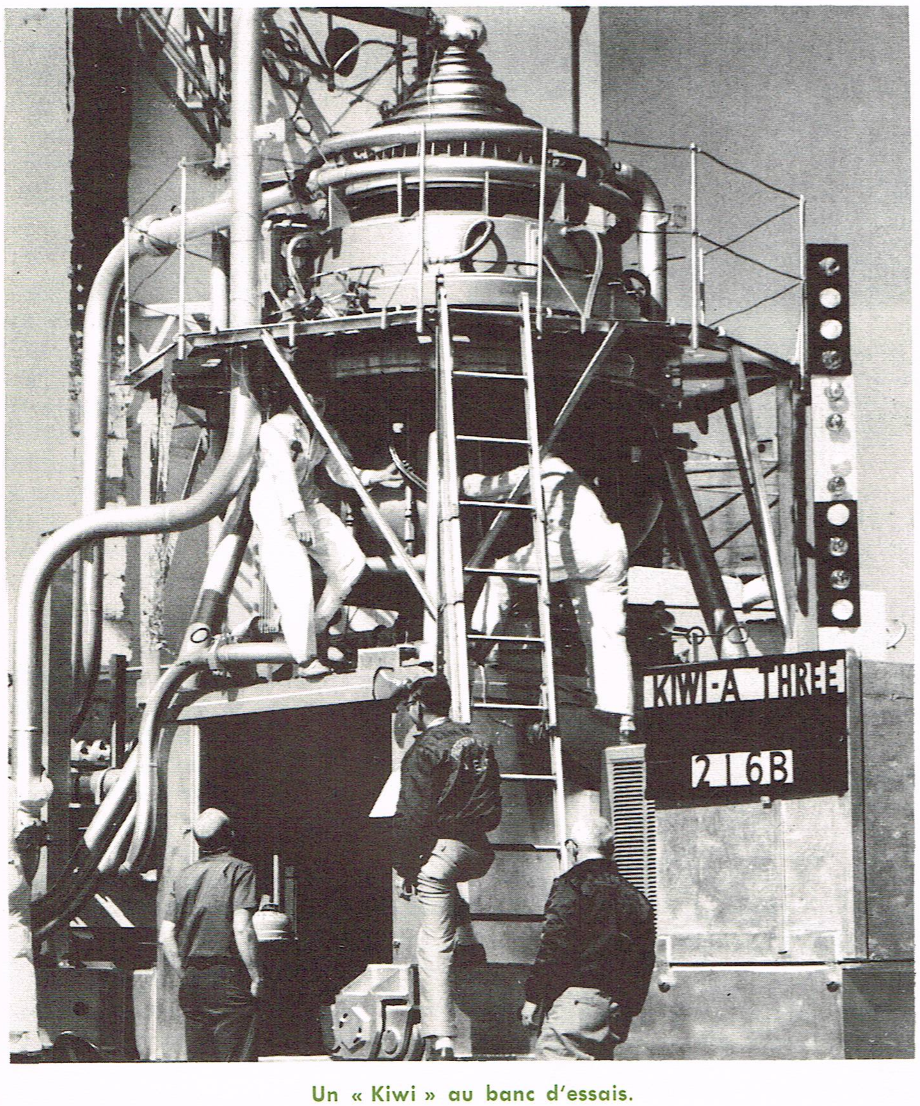
Dans tous les cas, il ne faut pas oublier que les systèmes nucléaires offrent des dangers particuliers dus au rejet dans la biosphère des produits radio-actifs. Ces risques sont toutefois réduits dans le cas très fréquent où le réacteur nucléaire n'est mis en route qu'au-delà d'une certaine altitude.

Les premières études relatives à la propulsion nucléaire dans l'espace furent axées sur le fonctionnement d'un réacteur nucléaire ayant l'hydrogène comme fluide réfrigérant. Elles sont conduites à Los Alamos sous l'égide du projet « Rover » qui, en décembre 1962, avait déjà absorbé un milliard de dollars, avec pour principal objectif un essai en vol en 1968. Plusieurs réacteurs de la série « Kiwi » furent réalisés et expérimentés au sol. Un essai capital eut lieu le 30 novembre 1962; le réacteur a été arrêté avant d'avoir atteint la pleine puissance en raison de lueurs apparues à la tuyère de sortie dont on n'a pas pu, ou voulu, expliquer la provenance.

Un réacteur « Kiwi » type B alimenté par l'hydrogène liquide est en cours de mise au point. Les sociétés Westinghouse et Aerojet développent un réacteur analogue — « Nerva » — destiné à l'équipement du deuxième étage d'une fusée « Saturn ».

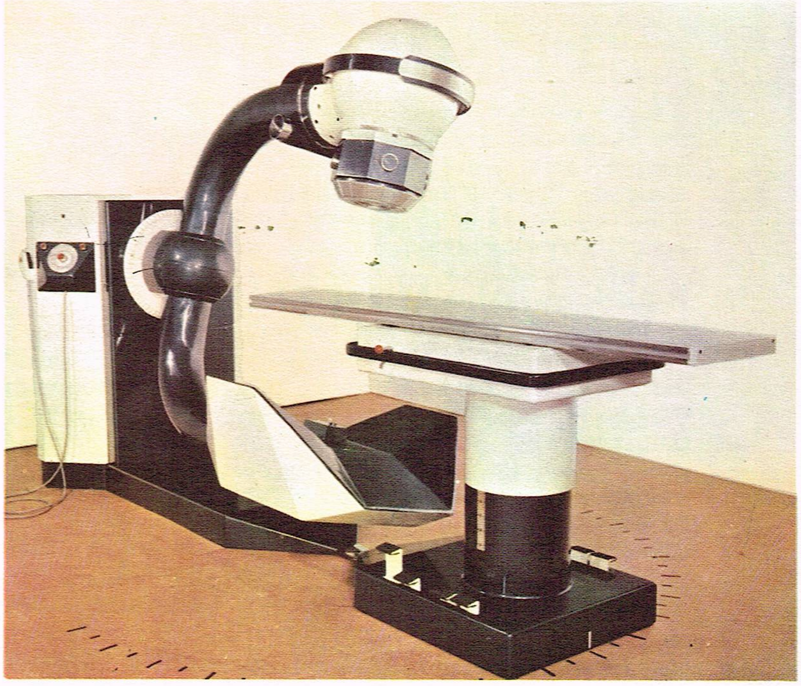
D'autres réacteurs, dits gazeux, utilisent un mélange formé d'hydrogène — stocké à 150 kg/cm² — et de matières fissiles en poudre; après passage dans une conduite en graphite, le mélange sort à la température de 10 000 °C, la puissance serait de l'ordre de 10 000 MW et le débit des gaz de l'ordre de 100 g/s.

Les avantages présentés par les réacteurs nucléaires « rapides » sont également envisagés.



Un « Kiwi » au banc d'essais.

Phot. U. S. I. S.



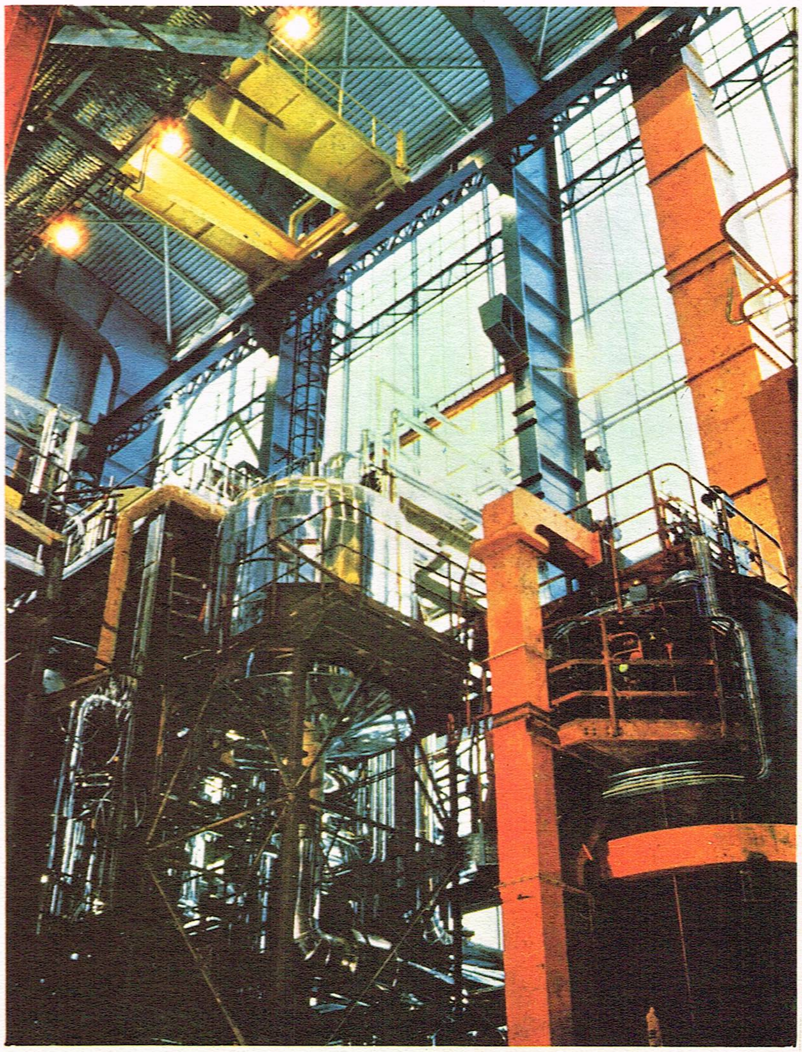
L'un des appareils de cobaltothérapie popularisés sous le nom de « bombe au cobalt ».

Phot. Philips.



La pile-piscine *Pégase*, du Centre d'études nucléaires de Cadarache, sert à essayer en vraie grandeur les éléments combustibles destinés aux réacteurs de puissance des centrales nucléaires.

Phot. C. E. A., Pierre Jahan.



Circuit expérimental pour l'étude, à Cadarache, du sodium fondu en tant que fluide caloporteur.

Phot. C. E. A., Pierre Jahan.

L'énergie nucléaire peut être utilisée à des fins moins spectaculaires que la propulsion; elle peut, après transformation, alimenter les appareils qui se trouvent à bord — comme ce fut le cas pour « Transit IV » lancé en juin 1961.

Pour cela, en 1956, l'AEC lança le programme SNAP — système nucléaire auxiliaire de puissance. Les premières réalisations utilisaient l'énergie thermique libérée par des radio-éléments artificiels, qui était transformée en énergie électrique en chauffant un thermo-élément; le « SNAP 3 » qui utilisait le polonium 210 fut présenté en janvier 1959 par le président Eisenhower : il développait 5 W. En choisissant d'autres corps radio-actifs, on arrive à quelques centaines de watts qui peuvent être débités sans arrêt pendant une dizaine d'années; la puissance est faible, mais ces dispositifs ont l'immense avantage de ne pas avoir de parties mobiles.

Aux Etats-Unis, ces générateurs sont utilisés à des fins non spatiales : alimentation de stations météo isolées, relais amplificateurs pour câbles sous-marins... Seul, leur prix de revient, encore très élevé, limite leur développement. **Mais pour les applications spatiales la sûreté de fonctionnement prime tout; peu importe le prix de la source d'énergie si elle peut multiplier par 10 ou par 50 la durée de vie utile des satellites.**

D'autres systèmes SNAP, où la source d'énergie est une véritable pile atomique, ont été étudiés. Ces réacteurs, qui utilisent presque toujours de l'uranium très enrichi, sont refroidis par des métaux fondus (potassium, sodium, rubidium, lithium) permettant d'atteindre des températures supérieures à 1 000 °C. Dans un échangeur ces fluides cèdent leur chaleur à un fluide moteur, qui est en général le mercure, qui à son tour entraîne une turbine à vapeur spéciale. Ces machines thermiques nécessitent une source froide (condenseur); au cours des vols spatiaux, le refroidissement ne pouvant être assuré que par rayonnement, il est donc nécessaire d'avoir des radiateurs de grande surface, même quand la température du condenseur est de l'ordre de 400 °C.

On arrive ainsi à des générateurs de quelques dizaines de kilowatts.

Certains projets en cours d'étude — « SNAP 50 » — envisagent de fournir une puissance électrique de quelques mégawatts; le but n'est plus alors d'alimenter des appareils de mesure, mais de faire de la propulsion électrique. L'énergie électrique permet d'ioniser un gaz, d'accélérer les ions par un champ électrique et de les éjecter avec une très grande vitesse. L'accroissement de vitesse ainsi communiqué à la fusée est toujours très faible, mais si l'on envisage des voyages de longue durée on peut gagner un temps énorme. Si une fusée quitte la zone d'attraction terrestre à la vitesse d'un kilomètre par seconde, et si elle se déplace à vitesse constante, il lui faudra, dans les meilleures conditions, quinze mois pour atteindre Vénus, alors que si le véhicule est soumis à une poussée égale au millième de son poids, un mois suffira.

Les résultats obtenus avec les SNAP ont jusqu'ici été très satisfaisants, et les programmes d'études et de réalisations se déroulent avec beaucoup de régularité.

De l'étude des programmes américains (ANP — PLUTO — ROVER — SNAP), qui semblent très complexes, se dégage une première conclusion : les efforts entrepris ne se relâchent pas, mais vont, au contraire, en s'amplifiant. L'inflation des programmes est si importante que les annulations même spectaculaires, comme celle du programme ANP, ne semblent pas avoir d'effets sensibles sur cette progression très impressionnante.

On ne peut pas s'empêcher de faire une comparaison avec l'évolution, lente et difficile, des centrales électriques nucléaires. Une des raisons est que les centrales nucléaires n'apportent pas au consommateur un produit nouveau; ce sont simplement des systèmes énergétiques qui essaient d'entrer en concurrence avec des systèmes déjà existants et très perfectionnés : le prix de revient de l'unité d'énergie produite a, dans ces conditions, une importance capitale et finalement c'est presque le seul paramètre qui guide le choix des exploitants. Pour les recherches spatiales, la situation est tout autre : sans source d'énergie nucléaire, il faut renoncer à explorer le système solaire; les hommes s'engagent donc dans des recherches extrêmement poussées, qui, étant donné l'importance des crédits nécessaires, seraient peut-être abordées avec moins d'ardeur si l'on avait en vue uniquement la production d'énergie électrique.

Ainsi se réalise la prophétie d'Esnault-Pelterie rappelée au début de ce chapitre et confirmée le 24 octobre 1961 au symposium international de Las Vegas par le Dr Seaborg, président de la Commission américaine de l'Energie atomique (AEC) : « Il existe deux grands domaines d'emploi pour la génération d'énergie à bord d'un véhicule spatial. Le premier est la propulsion, le second est l'alimentation de tous les instruments auxiliaires. Dans ces deux domaines, l'énergie nucléaire offre des avantages exceptionnels que l'AEC va exploiter en coopération avec les autres administrations nationales : l'industrie et les universités. L'un des problèmes essentiels est la propulsion électrique des engins spatiaux. Pour les missions lointaines, ce mode de propulsion paraît indispensable et on ne peut le réaliser qu'à partir de l'énergie nucléaire. »

l'énergie thermonucléaire de fusion

Reportons-nous à ce qui a été indiqué, par ailleurs, sur les réactions nucléaires et aux indications numériques sur les masses des noyaux légers, nous verrons qu'un grand nombre de réactions de fusion sont pensables. Nous en indiquons ci-après quelques-unes :

$D + D = T + p + 4{,}00$ MeV.
$D + D = n + He^3 + 3{,}28$ MeV.
$D + T = n + He^4 + 17{,}6$ MeV.
$He^3 + D = p + He^4 + 18{,}3$ MeV.
$Li^6 + D = He^4 + He^4 + 22{,}4$ MeV.

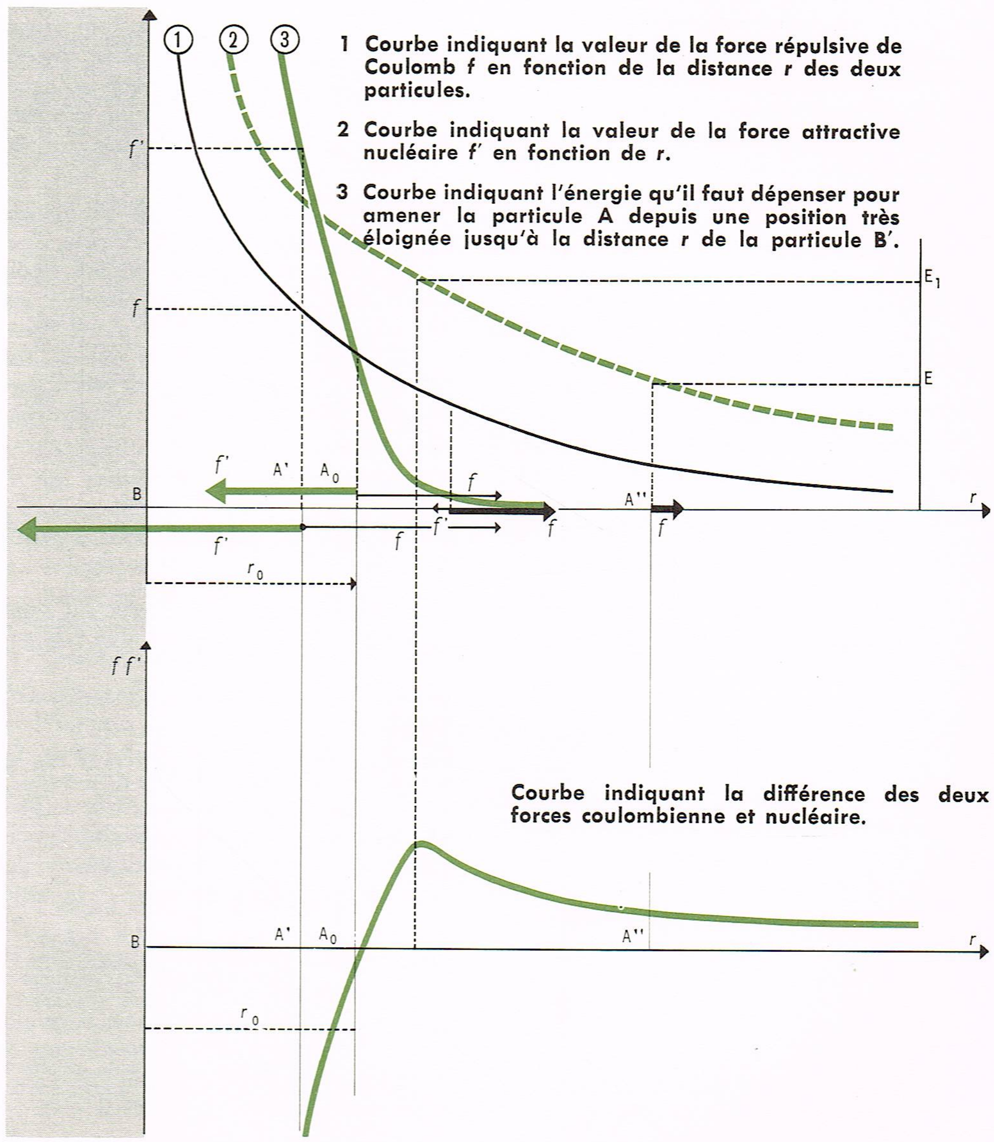
Les symboles D et T représentent le « deuton » et le « triton », c'est-à-dire les noyaux de l'hydrogène lourd (proton + un neutron) et de son isotope supérieur (proton + deux neutrons).

Pratiquement, seules les trois premières réactions présentent de l'intérêt, du fait que l'énergie libérée par nucléon mis en jeu est au moins aussi importante que dans la fission. De plus, et surtout, le deutérium est très abondant; il se trouve dans l'eau naturelle, dans la proportion de 1/6 000 de l'hydrogène léger, ce qui représente au total environ 5.10^{13} tonnes pour l'ensemble des océans. Le deutérium constitue le seul « combustible » terrestre pratiquement inépuisable, alors que, pour les autres, on commence à entrevoir le fond du sac.

un problème difficile : faire se rencontrer deux nucléons

Mais comment faut-il s'y prendre pour réaliser de telles réactions de fusion? L'expérience montre que les sections efficaces de ces réactions, même pour les trois premières, sont très faibles, beaucoup plus faibles que celles de la fission. De plus, dans le domaine des énergies intéressantes (nous verrons plus loin que c'est quelques dizaines de kilo-électron-volts, contrairement à ce qui se passe pour la fission), les sections augmentent avec l'énergie. Avant de tirer des conséquences de ces résultats expérimentaux, nous en expliquerons les principales causes.

Deux nucléons (D et D, ou D et T) ne peuvent « fusionner » que s'ils sont en contact. Il faut donc les rapprocher; pour cela une première difficulté surgit. Alors qu'aucune force n'empêche le neutron de se rapprocher du noyau d'uranium pour provoquer une fission, un noyau de deutérium A qui se rapproche d'un autre noyau B — deutérium ou tritium — est repoussé par la force électrostatique de Coulomb f suivant la formule : $f = k\frac{e^2}{r^2}$, où k est une constante qui dépend des unités choisies, e la charge des particules et r la distance qui les sépare. Tout se passe comme s'il existait entre A et B

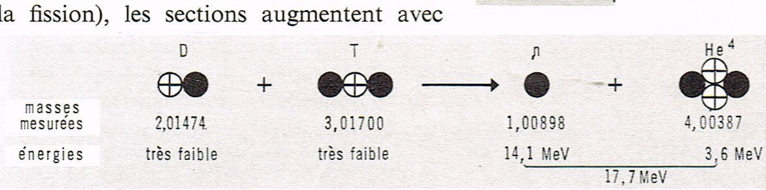


1 Courbe indiquant la valeur de la force répulsive de Coulomb *f* en fonction de la distance *r* des deux particules.

2 Courbe indiquant la valeur de la force attractive nucléaire *f'* en fonction de *r*.

3 Courbe indiquant l'énergie qu'il faut dépenser pour amener la particule A depuis une position très éloignée jusqu'à la distance *r* de la particule B'.

Courbe indiquant la différence des deux forces coulombienne et nucléaire.



L'ensemble des particules du deuxième membre a une masse plus faible que celui des particules du premier membre. La différence, 0,01889, permet, par la formule d'Einstein, de calculer l'énergie libérée.

un ressort travaillant à la compression, que A doit comprimer pour se rapprocher de B. Cette force augmente en même temps que r décroît. On comprend que, pour vaincre la résistance de cette force, le noyau doive effectuer un travail, qui sera d'autant plus important que la distance qui séparera les deux noyaux sera plus faible. Si l'on fournit au noyau A de l'énergie cinétique, il pourra l'utiliser pour vaincre la force de Coulomb, et la distance minimale r_m d'approche sera d'autant plus faible que l'énergie E sera plus forte; le produit Er_m est, en effet, presque constant et voisin de ke^2.

Si les forces de Coulomb étaient seules en jeu, il faudrait probablement renoncer à faire fusionner deux noyaux légers. Il existe heureusement d'autres forces attractives qui s'exercent entre deux nucléons; ce sont ces forces, dites « nucléaires », qui assurent la stabilité de l'échafaudage de neutrons et de protons qui constituent les noyaux de tous les corps. Ces forces sont pratiquement nulles, sauf quand la distance qui sépare les noyaux est très faible — disons de l'ordre de 10^{-12} cm. Elles augmentent quand r diminue, à une distance r_0 elles équilibrent l'effet répulsif des forces de Coulomb.

Il semble donc que si l'énergie du nucléon A est inférieure à E, il ne pourra pas y avoir rapprochement (donc pas de possibilité de fusion), alors que, dans l'hypothèse contraire, les deux noyaux s'agglomèrent pour former un élément plus lourd.

C'est au moins les conclusions que l'on pourrait tirer si les phénomènes que nous examinons étaient gouvernés par les lois de la mécanique rationnelle.

Or, aux échelles nucléaires, c'est une nouvelle mécanique : la mécanique ondulatoire, qui décrit plus correctement les phénomènes observés. Et les conclusions auxquelles on aboutit sont tout à fait différentes : quelle que soit la valeur de l'énergie E fournie au noyau, il peut fusionner. Cette fusion n'est pas certaine, mais elle a une probabilité bien définie de se réaliser, qui augmente, du reste, avec E — au moins dans la zone d'énergies envisagées (tout cela sera expliqué en détail dans le chapitre « la Matière, onde ou corpuscule », consacré à la mécanique ondulatoire).

Nous venons de définir des conditions nécessaires pour que deux deutons (ou un deuton et un triton) fusionnent. Mais quand ces deux particules arrivent en contact, il peut se produire d'autres réactions : choc élastique entre autres. De fait, cette dernière réaction est beaucoup plus fréquente que la fusion — cent fois plus environ.

D'autre part, il faut bien voir que l'énergie que l'on a fournie au deuton A peut être utilisée à autre chose qu'à « gravir la montagne de potentiel », et en fait, dans certains cas, cette énergie est presque totalement utilisée pour ioniser le milieu dans lequel se déplace le deuton.

Tout cela explique les très faibles sections efficaces. Pour l'ensemble des deux réactions DD, le maximum (de l'ordre de 0,1 barn) a lieu vers 750 keV. Mais à 60 keV, on est déjà à 10 p. 100 du maximum. Pour la réaction DT, le maximum a lieu vers 150 keV, et les 10 p. 100 sont atteints vers 40 keV.

Il faut retenir principalement les conclusions suivantes : pour contrebalancer les faibles sections efficaces, on doit opérer avec des faisceaux ayant une forte densité de noyaux et des énergies de l'ordre de quelques dizaines de kilo-électron-volts. On pourrait donc imaginer qu'avec deux accélérateurs on engendre deux faisceaux de deutons, ayant chacun une énergie de l'ordre de 60 keV, et qu'on les projette de front l'un contre l'autre. On aura bien des réactions de fusion, mais en nombre infime, car la densité des faisceaux produits par les accélérateurs actuels est extrêmement faible : il faudrait la multiplier par 10 000 ou 100 000, pour que la machine ait une densité de puissance par centimètre cube qui soit intéressante.

des énergies se chiffrant par centaines de millions de degrés

Il faut donc chercher dans une autre voie. Celle des hautes températures est la plus prometteuse.

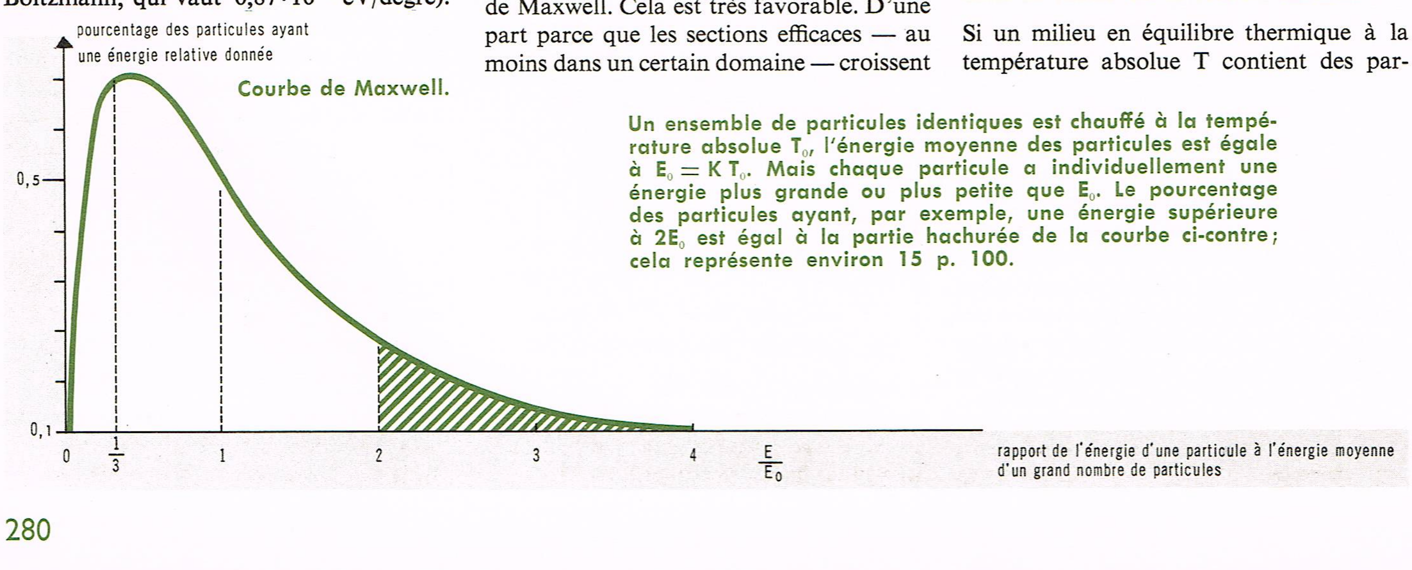
Rappelons que quand un gaz ou un liquide est chauffé à la température absolue T, l'énergie moyenne de ses molécules est égale à : $\mathbf{E = kT}$ (k est la constante de Boltzmann, qui vaut $0{,}87 \cdot 10^{-4}$ eV/degré).

Cette formule permet de repérer une température, aussi bien en °K qu'en énergie.

Mais si l'on examine individuellement chaque molécule, certaines auront une énergie beaucoup plus petite que E, d'autres une énergie beaucoup plus grande, la répartition ayant lieu selon une courbe de Maxwell. Cela est très favorable. D'une part parce que les sections efficaces — au moins dans un certain domaine — croissent rapidement avec l'énergie, d'autre part parce que, si nous chauffons un milieu contenant des deutons, cible et projectile seront en mouvement.

Un calcul détaillé montre qu'on gagne de ce seul fait un facteur voisin de quatre, les vitesses relatives intervenant seules dans la valeur de la section efficace.

Si un milieu en équilibre thermique à la température absolue T contient des par-



Courbe de Maxwell.

Un ensemble de particules identiques est chauffé à la température absolue T_0, l'énergie moyenne des particules est égale à $E_0 = KT_0$. Mais chaque particule a individuellement une énergie plus grande ou plus petite que E_0. Le pourcentage des particules ayant, par exemple, une énergie supérieure à $2E_0$ est égal à la partie hachurée de la courbe ci-contre; cela représente environ 15 p. 100.

ticules de deux espèces, avec des densités (nombre de noyaux par centimètre cube) respectives N_1 et N_2, on peut calculer les sections efficaces et le nombre des réactions Q par centimètre cube et par seconde, qui peut s'écrire

$$Q = N_1 N_2 g(T).$$

On constate que les réactions ne commencent à devenir abondantes qu'à partir d'une vingtaine de kilo-électron-volts; mais cela correspond à une « température » de 200 millions de degrés!

A ces températures, si on peut y arriver, toute la matière sera vaporisée et ionisée, et il est évident que les questions de température et de confinement seront capitales.

Par ailleurs, en tenant compte de l'énergie libérée par les réactions de fusion, on peut calculer la puissance volumique en fonction de la température T (ou de l'énergie) et de la densité de noyau (pour simplifier nous avons supposé $N_1 = N_2$). On a donc $W = \varepsilon N_1 N_2 g(T)$, ε étant l'énergie libérée par la fusion.

Il ne faut pas oublier que, finalement, il faudra extraire cette énergie. On ne peut donc envisager de trop fortes puissances volumiques, ni de trop faibles, car alors la machine cesserait de présenter tout intérêt économique.

Fixons-nous comme ordre de grandeur 100 watts par centimètre cube, chiffre atteint et même dépassé dans beaucoup d'autres machines. Cela conduit pour la réaction DT à des densités de noyaux de l'ordre de 10^{-4} de la densité normale et à des pressions de plusieurs centaines de milliers d'atmosphères pour 110 000 000 °K, température correspondant à 10 keV.

Nous voyons donc qu'il faut chauffer les deutons au-delà d'une certaine température pour avoir une puissance spécifique qui soit intéressante. Il est bien évident que ce milieu de deutons, chauffé à une telle température, aura tendance à se refroidir. Il faudra que l'énergie libérée par la fusion compense au moins ces pertes; les pertes par centimètre cube et par seconde sont données par une formule de la forme $W_1 = \varepsilon N_1 N_2 h(T)$, qui ne diffère de la formule précédente fournissant la puissance de fusion que par le facteur $h(T)$, qui remplace $g(T)$. Or, il se trouve que $h(T)$ et $g(T)$ croissent tous deux avec la température — ou l'énergie. Mais g croît beaucoup plus vite que T; il y aura donc une température critique T_c pour laquelle il y aura égalité entre les deux énergies. C'est cette température qu'il faudra dépasser pour que s'auto-entretienne la réaction.

Les pertes se produisent finalement par la surface du système, alors que l'énergie est libérée dans chaque élément de volume. Pour diminer l'importance relative des pertes, il faudra donc diminuer le rapport surface-volume, c'est-à-dire augmenter les dimensions du système : on arrive ainsi à la notion d'une taille critique.

La théorie montre que la fonction $h(T)$ qui permet de calculer les pertes est — toutes choses égales par ailleurs — proportionnelle au carré Z^2 du nombre Z de protons compris dans le noyau (nombre atomique). On a donc le plus grand intérêt à supprimer les éléments lourds et à avoir un gaz de deutons aussi pur que possible.

Il est aussi évident que la condition ($T > T_c$) devra être réalisée pendant un certain temps et dans un certain volume : il faut que les réactions aient le temps de s'amorcer.

Le libre parcours moyen λ de fusion d'un deuton animé d'une vitesse v étant : $\lambda = 1/N\sigma$, le temps de vie θ d'un neutron sera donc égal au temps qui lui est nécessaire pour parcourir cette distance :

$$\theta = \frac{\lambda}{v} = \frac{1}{Nv\sigma}.$$

Si la température n'est pas maintenue pendant un temps de l'ordre de θ, le deuton, quand il aura parcouru λ, se trouvera dans un milieu froid ($T < T_c$) et les réactions cesseront. Finalement, le milieu devra rester chaud pendant un temps τ supérieur à θ, ce qui peut s'écrire $N\tau > \frac{1}{v\sigma}$.

Pour la réaction DD, $N\tau$ doit être supérieur à 10^{16}, et pour la réaction DT à $3 \cdot 10^{14}$. T est de l'ordre de T_c, et la pression proportionnelle à N. Pour ne pas cumuler les difficultés imposées par les hautes températures à celles inhérentes aux hautes pressions, on a intérêt à diminer N (donc la pression) et augmenter T.

Nous voyons que, finalement, il y a trois problèmes à résoudre :

1. Chauffer au-delà d'une température critique : 50 millions de degrés pour la réaction DT, 400 millions pour une réaction DD;

2. Ce gaz chauffé à ces températures devra être confiné, c'est-à-dire rester dans l'espace prévu pour la réaction;

3. Ces conditions devront être maintenues suffisamment longtemps.

un problème ardu : le confinement de la matière

Le problème le plus ardu est celui du confinement (ce qui ne veut pas dire que celui du chauffage soit facile). On ne peut pas songer à un récipient matériel pour maintenir — même pendant un temps court, de l'ordre de la milliseconde — des gaz à des températures de l'ordre du million de degrés et aux pressions correspondantes : un tel récipient serait rapidement volatilisé; bien avant même, dès que la température serait suffisante pour que les parois de ce réservoir commencent à se vaporiser, des éléments lourds se mélangeraient aux deutons, les pertes augmenteraient très rapidement et la température ne pourrait être maintenue.

Il faut donc opposer aux deutons un « obstacle immatériel », tel un champ magnétique.

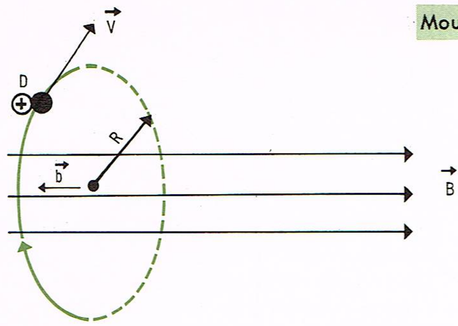
Comme toute particule chargée, un deuton animé d'une certaine vitesse décrit, sous l'action d'un champ magnétique uniforme perpendiculaire à sa vitesse, un cercle de rayon $R = l\,E/B$, formule dans laquelle l est une constante, E l'énergie du deuton et B l'intensité du champ. En prenant pour E 40 keV, on trouve pour le rayon :

4 cm pour B = 10 000 gauss,
4 mm pour B = 100 000 gauss.

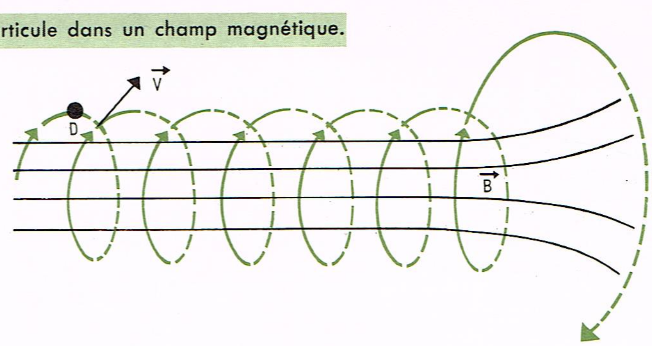
Il semble qu'on ait là un moyen efficace pour confiner; mais malheureusement les vitesses des particules ne sont pas toujours perpendiculaires aux lignes de force du champ : les deutons décrivent alors des trajectoires hélicoïdales et finissent par fuir. Pour éviter ces fuites, il faut trouver des configurations de champs fermés (en forme d'anneau) ou longitudinaux ayant aux deux extrémités des structures spéciales qui réfléchissent les particules comme le feraient des miroirs.

La particule qui tourne autour d'une ligne de force est assimilable à un courant électrique; les règles de l'électromagnétisme montrent qu'il se crée ainsi un champ opposé au champ initial $\vec{B}$. Tant qu'il y a une seule particule, le champ $\vec{B}$ n'est pratiquement pas modifié; mais quand il y en a « beaucoup », on comprend que

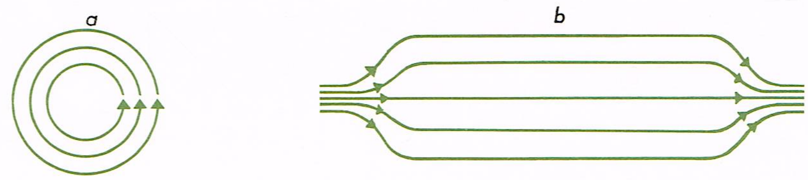
Mouvement d'une particule dans un champ magnétique.



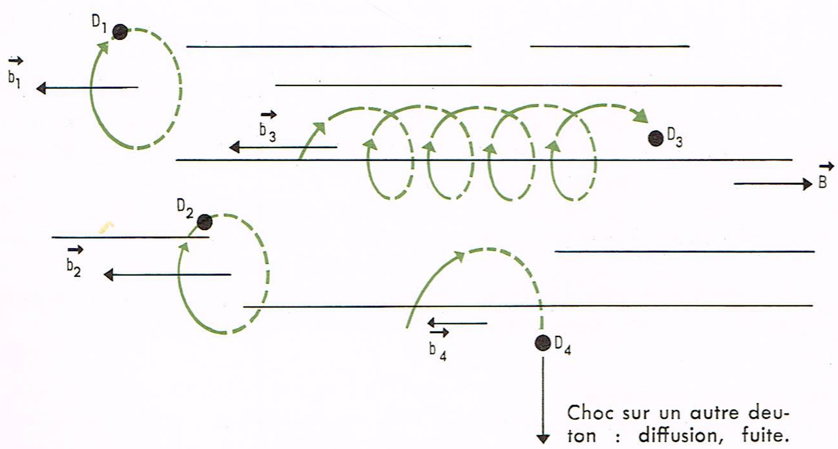
Si le champ est uniforme et si la vitesse est normale aux lignes de force, le deuton D décrit un cercle de rayon $R = k\frac{E}{B}$. Le déplacement de D équivaut à un courant qui crée un champ $\vec{b}$ opposé à $\vec{B}$.



Si la vitesse $\vec{V}$ est oblique, D décrit une hélice et se déplace le long des lignes de force et finit par fuir quand il arrive aux extrémités du champ. Pour essayer d'éviter ces fuites, on a deux groupes de solutions : *a*) champ torique ; *b*) tube de force fermé aux extrémités.

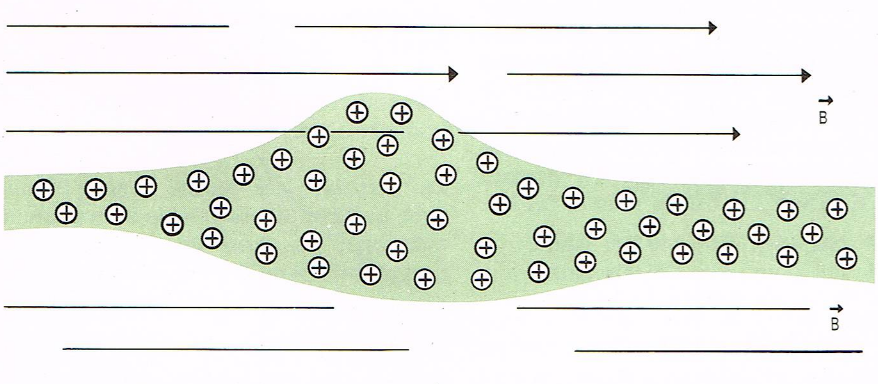


Mouvement de plusieurs particules dans un champ magnétique.

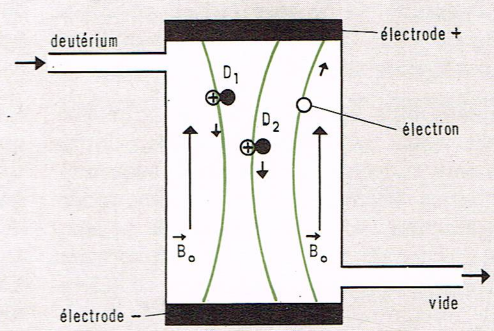


Choc sur un autre deuton : diffusion, fuite.

Les champs $\vec{b}$ s'additionnent ; quand il y a beaucoup de particules, leur somme est du même ordre que $\vec{B}$, il n'y a plus de confinement : le plasma « fuit » à travers le champ.

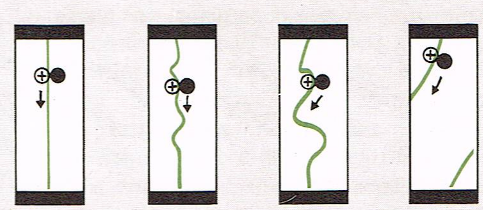


Striction linéaire.

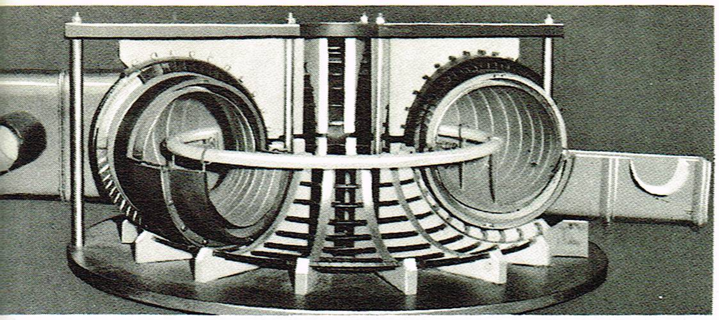


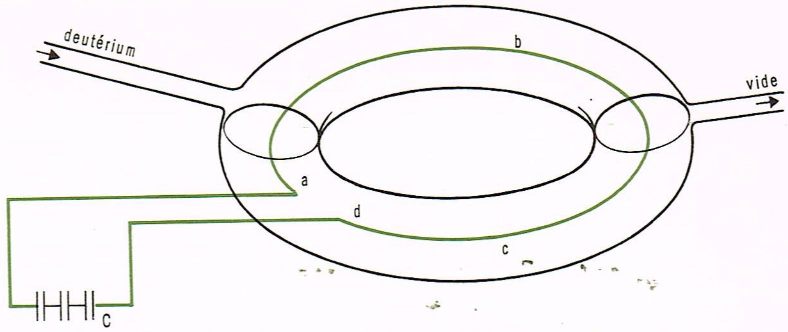
Les courants électriques constitués par le déplacement des deutons et des électrons s'attirent. Toutes les particules se concentrent au centre du récipient. La même décharge électrique produit à la fois le chauffage — effet Joule — et le confinement.

Ces décharges sont instables :

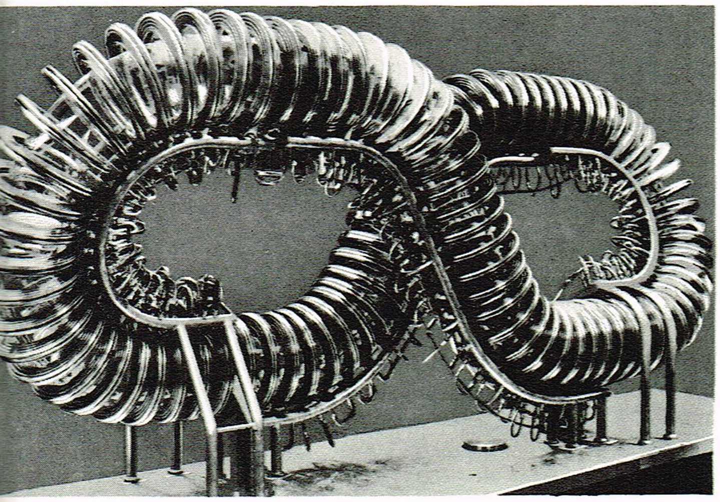


On diminue l'instabilité en ajoutant un champ longitudinal B_0 ; mais B_0 est toujours très inférieur au champ de confinement créé par le déplacement des deutons.



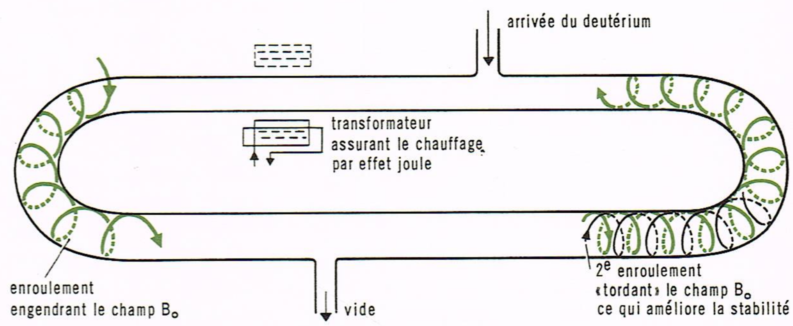


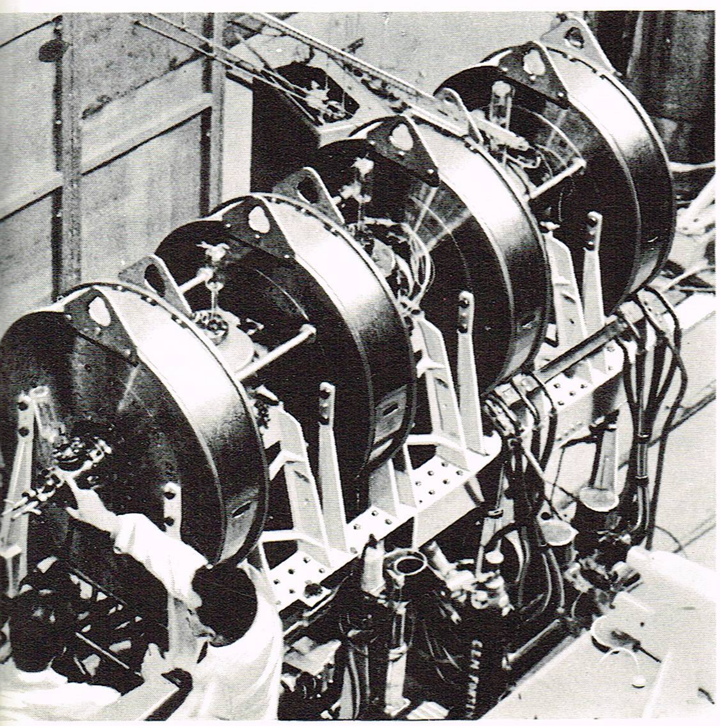
Maquette d'un appareil soviétique pour les recherches sur la **STRICTION TORIQUE.** Le tore est constitué par une matière isolante étanche. On décharge une batterie de condensateurs C dans la spire *a b c d* située à l'intérieur du tore. Cette spire constitue le primaire d'un transformateur, dont le secondaire est constitué par les deutons.



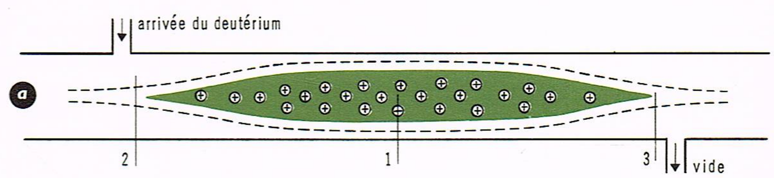
« Stellarator » américain, constitué par un tube fermé sur lui-même. Le confinement est assuré par le champ longitudinal B_0, qui est — contrairement à ce qui se passe dans la striction — très supérieur au champ radial créé par le déplacement des deutons.

Le chauffage est assuré par un dispositif indépendant de la création du champ de confinement.



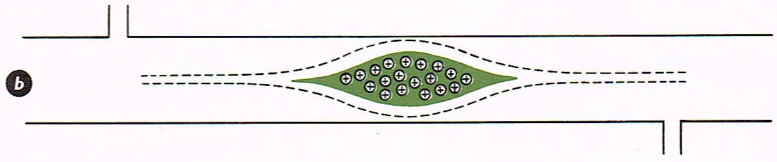


Bouteille magnétique au Centre de recherches nucléaires de Fontenay-aux-Roses.



Un premier groupe de bobines crée un champ ayant la configuration ci-dessus; il est minimal en 1 et maximal en 2 et 3. Avec une telle disposition, une particule, qui arrive en 3 venant du centre, est ralentie, stoppée et renvoyée vers le centre.

Un deuxième groupe de bobines crée un champ variable qui se superpose au champ ci-dessus; il y a un brassage du plasma, qui oscille entre les positions *a* et *b* et qui s'échauffe en même temps.



Phot. Larousse, U. S. I. S., Majo.

le champ peut être sensiblement réduit, et même annulé; il ne peut donc plus confiner. On peut préciser l'adverbe « beaucoup » et calculer quel est le nombre n de particules par centimètre cube qui annulent le champ $\vec{B}$. On trouve : $8\pi n E = B^2$, E étant l'énergie moyenne des particules ($E = kT$).

La valeur de E étant imposée par la température critique, on en déduit que n varie comme B^2, carré du champ. En pratique, cependant, n doit être sensiblement inférieur aux minimums théoriques ainsi calculés :

$$n < 10^{14} \quad (B = 10\,000 \text{ gauss}).$$

$$n < 10^{16} \quad (B = 100\,000 \text{ gauss}).$$

Mais il ne faut pas oublier qu'il y a toujours des chocs de particules entre elles; les particules dévient de leurs trajectoires théoriques et peuvent même finir par quitter le système (diffusion). La théorie exacte de ce phénomène est très compliquée; en gros, l'intensité du phénomène varie comme 1/B.

Enfin, il y a encore une difficulté, qui est peut-être la plus grave : jusqu'ici on pouvait concevoir l'ensemble des particules comme contenues à l'intérieur d'un cylindre fictif; à l'extérieur de ce cylindre régnait le champ B, à l'intérieur la valeur de ce champ était diminuée des champs dus aux courants électriques engendrés par le mouvement des particules. La limite entre ces deux zones n'est pas fixe, l'expérience montre qu'elle est soumise à des mouvements plus ou moins violents et que très souvent le phénomène n'est pas stable, c'est-à-dire que les particules ne restent pas confinées. Les conditions de stabilité ou d'instabilité ne peuvent s'exprimer par une règle simple; pour avoir un régime stable, on choisit actuellement des densités de particules de l'ordre du dixième de celles qui sont indiquées par la formule, c'est-à-dire qu'on prendra

$$8\pi n E = B^2/10.$$

Le problème du confinement étant supposé résolu, il faut maintenant s'occuper de « chauffer » les particules. Plusieurs techniques sont utilisées :

1. Chauffage par effet Joule; ce procédé donne de bons résultats jusqu'à quelques millions de degrés;

2. Pour chauffer au-delà, on fait osciller rapidement un champ magnétique, les oscillations mettent en mouvement deutons et électrons, les agitent dans tous les sens; ce qui équivaut à une augmentation de température. Ce procédé s'appelle le « pompage magnétique ».

demain : des centrales thermonucléaires ?

Faisons un saut dans le futur et supposons que toutes les difficultés signalées (confinement, chauffage, maintenance) aient été résolues, et cherchons comment pourrait fonctionner une centrale industrielle fournissant de l'électricité d'origine thermonucléaire.

Ecrivons à nouveau les principales réactions de fusion en répartissant à chaque particule l'énergie totale libérée :

$$D+D = (T + 1{,}0\ MeV) + (p + 3{,}0\ MeV)$$

$$D+D = (He^3 + 0{,}8\ MeV) + (n + 2{,}48\ MeV)$$

$$D+T = (He^4 + 3{,}6\ MeV) + (n + 14{,}0\ MeV).$$

Nous remarquons que dans l'ensemble des deux réactions D D, un tiers de l'énergie libérée est communiquée aux neutrons; dans le cas de la réaction DT, ce sont les 4/5 qui sont pris par les particules neutres.

Or, les neutrons, insensibles aux champs électriques ou magnétiques, peuvent facilement quitter le système et perdre leur énergie cinétique dans un modérateur analogue à ceux qui sont utilisés dans les réacteurs à fission : un fluide réfrigérant enlève l'énergie cédée par les neutrons au modérateur, et peut ensuite actionner des turbines.

Il serait aussi possible d'utiliser les neutrons en les absorbant dans du lithium selon la réaction

$$n'_o + Li_3^6 = He_2^4 + T_1^3 + 4{,}6\ MeV;$$

le tritium ainsi formé pourrait être séparé chimiquement et utilisé comme combustible pour les réactions D T.

On pourrait ainsi imaginer des machines fonctionnant par impulsions successives.

A chaque impulsion, des réactions de fusion se déclenchent, il y a une brusque élévation de température et des déplacements de particules chargées, c'est-à-dire des courants électriques. Ces courants peuvent produire, par induction dans des spires de fils conducteurs disposés autour de la machine, un courant pulsé facilement transformable en courant continu ou alternatif. **Mais un grand nombre de problèmes, tous difficiles, sont encore à résoudre avant de penser à l'organisation des futures centrales thermonucléaires, qui posent aux ingénieurs un des problèmes les plus excitants qui aient été jusqu'ici abordés par les hommes.**

les programmes civils

Au cours de l'année 1966, le monde occidental aura dépensé quelque 40 milliards de nouveaux francs pour l'énergie atomique; il est plus difficile de connaître exactement ce qui a été dépensé depuis l'origine : le chiffre ci-dessus doit être multiplié au moins par dix, peut-être par vingt.

Nous allons essayer d'indiquer rapidement comment et pourquoi des sommes aussi importantes ont été dépensées.

Grande-Bretagne

En 1955, l'Angleterre, qui, traditionnellement, était exportatrice de charbon, devint importatrice; depuis quelques années, les pétroles iraniens ont été nationalisés.

Dans un tel climat, le spectre de la pénurie d'énergie surgit. Nécessité faisant loi, pour éviter une hémorragie de devises, pour ne plus courir de risques quant à l'approvisionnement du combustible, le gouvernement britannique, dans son Livre blanc de 1955, décide la construction de centrales nucléaires d'une puissance totale de 2 000 MWE (1), qui devront fonctionner en 1966.

Les conclusions du fameux rapport des trois sages et le démarrage satisfaisant de Calder-Hall, qui fut inauguré par la reine le 17 octobre 1956, incitent le gouvernement à publier en mars 1957, un second Livre blanc; le programme initial est augmenté et accéléré : 6 000 MWE d'origine nucléaire, le quart de la puissance installée, doivent être produits en 1965.

En 1958 (un an après le second Livre blanc), le charbon et le pétrole deviennent surabondants, leurs prix diminuent; le prix de revient des centrales nucléaires ne baisse pas, malgré l'augmentation des puissances unitaires, mais celui des centrales classiques passe de £ 70 à £ 40 le kWE installé. Le programme des centrales nucléaires est ramené à 5 000 MWE. Ainsi petit à petit, la belle confiance qu'exprimait l'énorme programme britannique s'effrite, et en juillet 1960, le programme fut à nouveau modifié et ramené à 3 000 MW installés en 1966 et à 5 000 MW en 1968.

Toutes ces modifications de programmes ne se déroulèrent pas sans heurts, et de sévères critiques, tant sur le plan technique que sur le plan industriel, furent adressées à l'U.K.A.E.A. (United Kingdom Atomic Energy Authority) et aux responsables du programme.

Néanmoins, le programme de juillet 1960 s'exécute dans les délais prévus; aujourd'hui toutes les centrales en service fonctionnent parfaitement et ont reçu la sanction d'une expérimentation industrielle. Petit à petit, les problèmes de fabrication sont résolus et le coût des investissements diminue d'une manière spectaculaire. En 1964, un nouveau plan, qui prévoit l'installation de 5 000 MWE, puis celle de 8 000 MWE d'origine nucléaire entre 1970 et 1975, étant adopté, on peut lancer un appel d'offre où sont mis en concurrence les réacteurs américains à eau légère et les réacteurs anglais. L'AGR l'emporte; les centrales de ce type qui utilisent de l'uranium légèrement enrichi se substitueront désormais aux réacteurs de la filière Magnox. Cette décision a une importance capitale; d'une part, le nucléaire est considéré maintenant comme compétitif, d'autre part l'industrie anglaise, qui a montré qu'elle était capable de résister aux géants américains, reprend confiance.

Par ailleurs, des efforts très importants sont consacrés au réacteur *SGHWR* et au breeder rapide — *Dounreay* —; les crédits consacrés à cette dernière filière sont deux fois plus importants que ceux qui sont attribués à l'AGR.

Afin de pouvoir alimenter en uranium faiblement enrichi les réacteurs AGR, l'usine de Capenhurst, qui fabriquait de l'uranium très enrichi pour les besoins militaires, est en cours de modification, afin de fournir uniquement de l'uranium à faible taux d'enrichissement pour les besoins civils.

France

Comme la Grande-Bretagne, la France a commencé par s'engager dans la voie des centrales U naturel, graphite, CO_2.

On a construit successivement à Marcoule les trois piles plutonigènes : G1, G2 et G3, qui représentent une puissance électrique totale supérieure à 70 MW, et qui ont divergé respectivement en 1956, 1958 et 1959.

Les résultats obtenus avec ces piles prototypes ont permis à l'E. D. F. de lancer son programme nucléaire, dont la première tranche comprend E. D. F. 1 (70 MW), E. D. F. 2 (200 MW), E. D. F. 3 (480 MW). E. D. F. 4, en cours de construction, de caractéristiques techniques identiques à E. D. F. 3, sera compétitive avec le classique français. D'autres étapes techniques sont en cours; Bugey I — 500 MWE —, avec ses éléments annulaires, sera une étape intermédiaire avant les réacteurs de 1 200 MWE.

Le Ve Plan prévoit qu'au cours des cinq années on va installer une puissance nucléaire comprise entre 2 500 et 4 000 MWE.

Un réacteur à eau légère pressurisée (type U. S. A.), construit en collaboration avec l'industrie belge et dans le cadre des accords Euratom, fonctionne actuellement dans les Ardennes, à Chooz.

Un surgénérateur rapide : « Rapsodie », a divergé début 1957; un réacteur de technique strictement française vient d'être vendu en Espagne; à Pierrelatte, une usine de séparation isotopique fournit depuis plusieurs années de l'uranium faiblement enrichi et, depuis quelques mois, de l'uranium fortement enrichi.

Si les réalisations françaises semblent moins importantes que celles des Britanniques, la politique menée a été empreinte d'une grande sagesse, car jusqu'ici l'évolution technique se faisait avec une telle rapidité qu'on n'a pas estimé rationnel de se lancer dans des réalisations industrielles importantes avant d'être sûr des qualités du réacteur choisi.

États-Unis d'Amérique

La politique des États-Unis fut fondamentalement différente de celle de la Grande-Bretagne.

D'une part, aucune pénurie d'énergie n'était en perspective; d'autre part, les États-Unis disposant de puissantes usines de séparation isotopique construites pour les besoins militaires, l'U enrichi existait en quantités importantes, et les études furent d'abord axées sur les réacteurs modérés à l'eau légère.

Par ailleurs, le gouvernement ayant consacré d'importants crédits pour le dévelop-

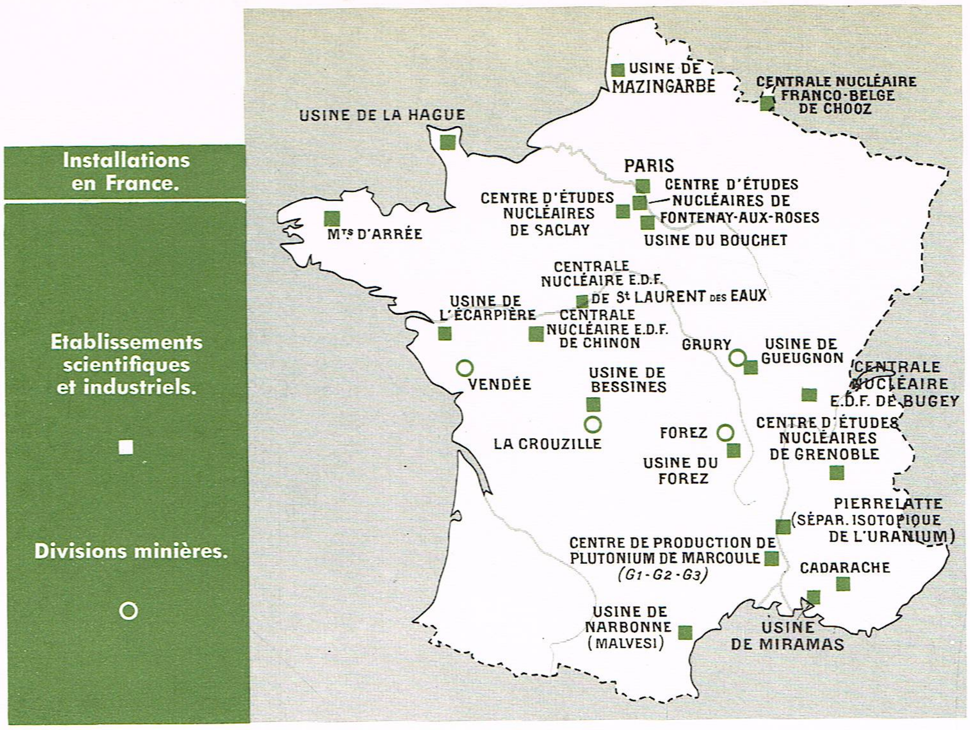
(1) Sigle du « mégawatt électrique » qui, dans les centrales nucléaires, mesure la puissance des alternateurs, alors que le « mégawatt thermique », employé par analogie, sert à exprimer l'équivalence en unités électriques de la chaleur dégagée par la fission au cœur du réacteur.



160° Est de Gr.
180°
160°
140° Ouest de Greenwich
120°
100°
60°
40°
20°
GRAND LAC DE L'OURS
LAC ATHABASCA
HANFORD
HUMBOLDT BAY
ARCO
ELK RIVER
ARGONNE
BLIND RIVER
CHALK RIVER
YANKEE
SIOUX FALLS
BERKELEY
NEVADA 1951
DRESDEN
ROCHESTER
INDIAN POINT
COLORADO
HALLAM
PIQUA
BROOKHAVEN
SANDIA
SHIPPING PORT
LOS ANGELES
LOS ALAMOS
ALAMOGORDO 16-VII-1945
OAK RIDGE
SAVANNAH
GUADALUPE
PADUCAH
OCÉAN
PACIFIQUE
OCÉAN
ATLANTIQUE
Tropique du cancer
I. JOHNSTON 1958
I. CHRISTMAS 1952
MURUROA
COLOMBIE
PARAIBA
MINAS GERAIS
ESPIRITO SANTO
Tropique du capricorne
Equateur
Cercle polaire arctique
HALDEN
GEESTHACHT
GOTTINGEN
AMSTERDAM
PETTEN
MOLL
DELFT
EUROCHEMIC
ARNHEM
JULICH
KARLSRUHE
BÂLE
LUCENS
ZURICH
ISPRA
BILBAO
SILVARES
SALUGGIA
MONCLOA
ANDUJAR
DOUNREAY
HUNTERSTON
CHAPEL CROSS
WINDSCALE
CALDER HALL
SPRINGFIELD
CULCHETH
CAPENHURST
RISLEY
TRAWSFYNYDD
SIZEWELL
OLDBURY
BRADWELL
BERKELEY
DUNGENESS
HARWELL
CORNOUAILLES
HINKLEY POINT
I. DE WIGHT
WINFRITH HEATH
OCÉAN
GLACI
Cercle
120° Ouest de Greenwich

PLANISPHÈRE DE L'ÉNERGIE NUCLÉAIRE





pement des sous-marins à propulsion nucléaire, plusieurs grandes sociétés, entre autres la General Electric Company et Westinghouse, acquirent d'énormes connaissances techniques, qui furent immédiatement utilisées pour construire des centrales nucléaires équipées de réacteurs P. W. R. et B. W. R.

En 1959, M. Mac Cone, président de l'AEC, définit un programme de 2,5 milliards de $ pour perfectionner les deux types de réacteurs ci-dessus et étudier les autres filières. Ce plan comporte aussi une assistance aux pays amis.

Ce grand effort est justifié aux yeux des Américains par le fait qu'ils espèrent rester le leader du monde occidental en matière atomique et arriver à mettre au point l'avènement des surgénérateurs.

Où en sont les États-Unis aujourd'hui ?

Quatre réacteurs de plus de 100 MWE sont en service, l'un d'eux depuis 1957; seize centrales, dont la puissance s'échelonne entre 500 et 1 000 MWE, sont en construction; plus d'une trentaine de projets (de 450 à 1 000 MWE) sont en cours de discussion entre les constructeurs de centrales et les producteurs d'électricité.

Au début de la mise en service des quatre réacteurs cités, il y a bien eu quelques incidents, mais aucun n'a pu mettre en cause la conception des réacteurs; il y a eu aussi quelques découragements parmi ceux qui croyaient à un démarrage foudroyant de l'énergie nucléaire, beaucoup auraient voulu être aidés plus efficacement par le gouvernement. Malgré tout cela, des efforts considérables furent faits, et, aujourd'hui, personne n'ose plus contester la compétitivité de l'énergie atomique.

Ce succès est principalement l'œuvre de deux géants : General Electric et Westinghouse, qui se sont respectivement spécialisés dans les *B. W. R.* et les *P. W. R.*

Malgré des fausses notes, malgré certaines luttes entre les dirigeants, il faut reconnaître que le gouvernement et l'industrie ont su imposer un rythme impressionnant au développement de cette nouvelle source d'énergie et qu'une œuvre immense est en train de se réaliser pour promouvoir l'utilisation de l'énergie atomique à des fins pacifiques.

U. R. S. S.

Probablement parce que ce pays dispose de ressources considérables d'énergie d'origine fossile, son programme de centrales nucléaires est moins important que ceux des pays que nous venons d'examiner.

Il faut toutefois rappeler que le premier réacteur qui a fourni des quantités notables d'électricité a été construit en 1954 — réacteur *AM 1* de 5 MWE; il était modéré au graphite et refroidi par de l'eau sous pression. Il a été suivi par un réacteur de 100 MWE, qui a fonctionné en 1963, et par un réacteur de 200 MWE, actuellement en cours de construction.

Le site de Novovoroulj a été équipé en 1964 d'un réacteur *P. W. R.* de 210 MWE; un deuxième réacteur de 365 MWE est en cours de construction; un réacteur bouillant, le *BK 50*, a divergé en 1965, et la voie des surgénérateurs est étudiée avec célérité : surgénérateur *BN 350*.

Canada

Jusqu'en 1964, le Canada a fait un effort continu et obstiné sur les réacteurs à eau lourde pressurisée. *NPD*, de 20 MWE, a été un succès; *CANDU*, de 200 MWE, démarrera en 1967.

Sur le site de Pickering (Toronto), on a prévu huit unités type *CANDU* de 500 MWE à installer dans les huit années à venir; de plus, le Canada doit exporter en Inde trois réacteurs de 200 MWE.

Très récemment, le Canada s'est occupé d'un réacteur un peu différent : modération à l'eau lourde et refroidissement par l'eau ordinaire bouillante : projet Hydro Québec.

Autres pays

Depuis dix ans, la Suède a fait un effort pour développer une filière originale — pressurisé à eau lourde —, mais des difficultés ont dû naître, car en 1965 la Suède a acheté à la General Electric un réacteur de 400 MWE bouillant.

L'Allemagne, l'Italie, le Japon, l'Inde, l'Espagne et la Belgique ont commencé par acheter des réacteurs étrangers.

L'Italie a expérimenté trois filières : un réacteur anglais (Latina), un *B. W. R.* (Garigliano) et un *P. W. R.* (Trino).

Indépendamment d'une demi-douzaine de petits réacteurs expérimentaux, l'Allemagne a actuellement en construction trois centrales de 250 MWE (*P. W. R.* et *B. W. R.*), d'autres sont en projet, et un effort de recherche important commence à se faire : surchauffe nucléaire, réacteurs modérés à l'eau lourde, surgénérateurs, etc.

Le Japon a acheté une centrale britannique en 1959 — avec graphite français —, Tokaï-*Mura*, de 165 MWE, a commandé récemment un bouillant de 325 MWE, un pressurisé de 250 MWE, et lance la construction d'un cargo nucléaire.

L'Inde a acheté à la General Electric un bouillant de 380 MWE, qui doit fonctionner en 1967, et 3 réacteurs de 200 MWE type *CANDU* au Canada.

L'Espagne a commandé un réacteur à la France type *E. D. F. 4* et deux aux Etats-Unis. La Finlande, le Danemark, le Mexique, la Nouvelle-Zélande, le Pakistan mettent au concours la construction de centrales.

les armes atomiques

L'histoire fait remonter à plusieurs siècles avant notre ère la mise en service de l'arc dans les armées : « Honni soit le premier qui fut archer. C'était un lâche, il n'osait approcher. »

Ce n'est pas d'hier que l'homme cherche des moyens pour augmenter la distance à laquelle il était possible de tuer et de détruire. Les progrès dans cette voie furent rapidement limités, car les flèches lancées par les arcs ou les arbalètes pouvaient emmagasiner seulement l'énergie que l'homme était capable de développer pendant un temps relativement court. Néanmoins, pendant près de deux millénaires, cette arme fut considérée comme très efficace et il fallut attendre 1346 pour voir, à la bataille de Crécy, des engins utilisant de l'énergie qui n'était pas tirée directement du travail musculaire que l'homme peut fournir.

1346 marque l'an I de l'ère du T. N. T. (trinitrotoluène : c'est encore l'explosif chimique le plus puissant que l'homme sache fabriquer); son déclin apparut six cents ans plus tard à Hiroshima : peu à peu le T. N. T. cède la place aux explosifs nucléaires.

Quelles sont les différences profondes qui distinguent un explosif nucléaire d'un explosif classique : T. N. T. par exemple? Avant de répondre à cette question, il faut préciser que les principaux explosifs sont constitués par des carbures d'hydrogène nitrés, qui s'oxydent en utilisant soit l'oxygène de l'air, soit l'oxygène contenu éventuellement dans leurs molécules.

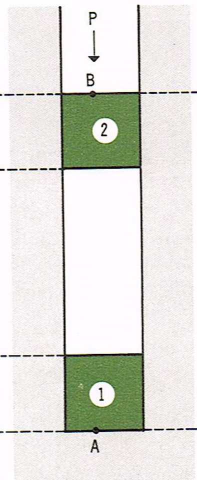
En étudiant ces réactions, on constate que l'énergie libérée par une réaction explosive (rapportée à la masse en jeu) est sensiblement inférieure (de l'ordre de dix fois) à celle qui est libérée par la combustion de la même masse de carbone. Quel est donc le phénomène qui confère autant de puissance aux explosifs? Leur efficacité est pratiquement due uniquement à leur très grande vitesse de réaction chimique, qui fait qu'en un temps de l'ordre du dix-millième de seconde, une masse très appréciable de matière est portée à une température de 3 000 à 4 000 °C. Comme les produits de la réaction sont gazeux et qu'ils sont — au moins pendant un certain temps — confinés par les parois du projectile, la pression croît jusqu'à des valeurs de l'ordre de 100 000 atmosphères. Finalement, les projectiles explosifs ne sont pas des systèmes qui développent beaucoup d'énergie, mais des engins qui créent de très grandes puissances pendant un temps court.

On voit alors immédiatement l'intérêt d'utiliser des réactions nucléaires à la place des réactions chimiques : d'une part, à égalité de masse mise en jeu, l'énergie libérée est, en effet, environ un million de fois plus importante; d'autre part — comme nous le verrons plus loin —, cette libération s'effectue dans un temps de l'ordre de la microseconde, c'est-à-dire cent fois plus faible que dans le cas d'un explosif classique. Cette performance n'est toutefois atteinte que quand on utilise des fissions provoquées par des neutrons rapides d'énergie 2 MeV; avec des neutrons thermiques, le temps d'explosion ne serait pas beaucoup plus court qu'avec les explosifs conventionnels; le « bénéfice » ne porterait que sur l'énergie libérée, beaucoup plus grande.

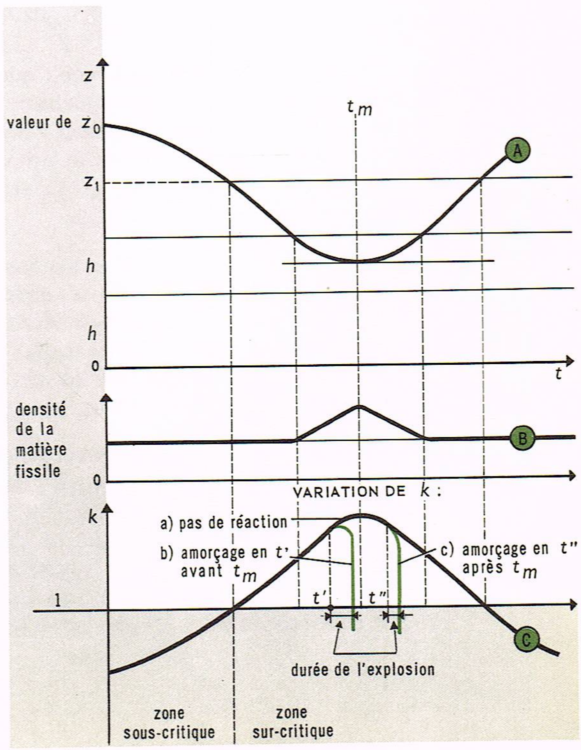
Comment peut-on organiser des armes atomiques? Nous avons déjà vu qu'il existe deux processus nucléaires permettant de libérer de l'énergie : la fission et la fusion; ils ont donné naissance à deux types de bombes, appelées respectivement « bombe A » et « bombe H ».

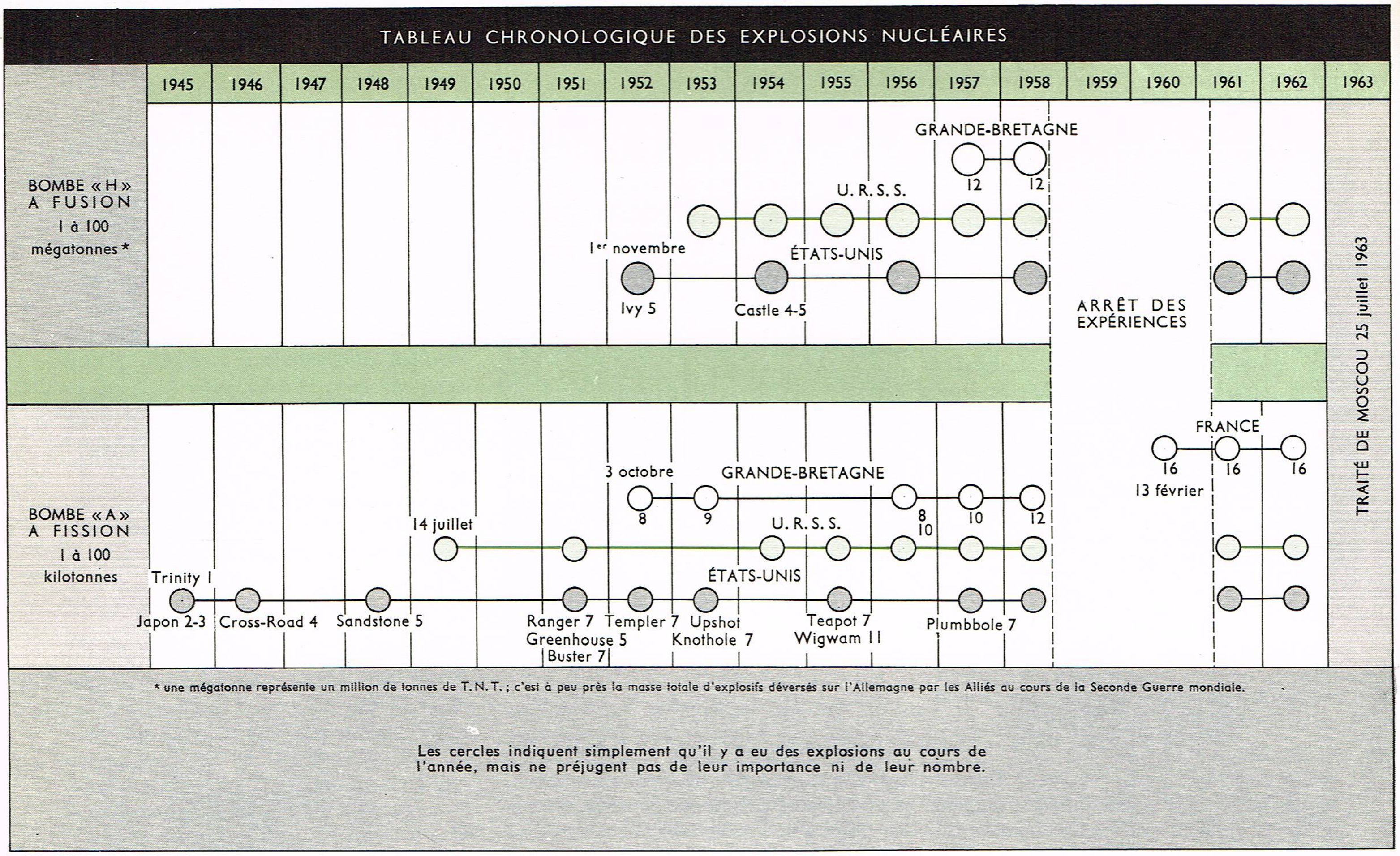
l'ancêtre : la bombe A

Occupons-nous d'abord de la première, qui fait exclusivement appel à la fission. Il est d'abord évident que, pour des raisons de sécurité, le système fissile qui constitue une arme doit être sous-critique quand celle-ci est au repos et ne devenir sur-critique qu'au moment du fonctionnement. On doit donc avoir recours à un système dont la géométrie se modifie. Un des plus simples est constitué par deux cylindres de matière fissile, égaux, sous-critiques quand ils sont pris isolément et sur-critiques quand ils sont rassemblés.



Imaginons que ces deux cylindres puissent glisser dans un tube cylindrique, à la manière d'obus dans l'âme d'un canon. Nous supposerons l'un fixe, l'autre mobile. La position relative de l'un par rapport à l'autre est définie par la distance z. Au repos, z est suffisamment grand pour que l'ensemble soit sous-critique. Une explosion conventionnelle précipite le cylindre mobile sur le fixe. Que va-t-il se passer? Si l'on ne tient pas compte des réactions nucléaires, on peut décrire le phénomène par la courbe qui exprime les variations de z en fonction du temps : la vitesse de la masse mobile va augmenter, jusqu'à ce que les deux masses soient en contact ($z = 2h$). A partir de ce moment, la vitesse diminuera — les deux masses seront comprimées et augmenteront de densité —, puis s'annulera; on aura ensuite le phénomène inverse : les deux blocs comprimés se détendront et s'éloigneront, au moins si la pression P due à l'explosion initiale a diminué ou a été supprimée.





Le système étant parfaitement défini par z et par la composition des deux masses, on peut calculer en fonction de z le facteur de multiplication efficace : il est inférieur à l'unité pour z_o, valeur initiale de z, égal à 1 pour une valeur z_l, comprise entre z_o et $2h$, et supérieur à 1 au-delà.

Pour qu'une suite de réactions en chaîne se produise, il faut qu'il y ait au moins un neutron qui vienne déclencher la première fission, pendant que le système est surcritique.

La chaîne étant amorcée, on conçoit que l'énergie libérée par les réactions nucléaires va s'opposer au rapprochement des masses si l'amorçage a lieu avant le temps moyen t_m, et au contraire accélérera l'éloignement s'il a lieu après t_m.

Il est évident que le nombre de fissions, c'est-à-dire l'énergie libérée, sera d'autant plus important que les réactions se dérouleront dans la zone voisine de k maximal : on aura donc intérêt à déclencher les réactions nucléaires avant t_m plutôt qu'après.

Pour arriver à libérer une puissance de 20 kt (c'est-à-dire 20 kilotonnes ou 20 000 t de T. N. T., ce qui correspond à l'énergie libérée par un kilogramme de matière fissile, dont les noyaux auraient tous subi une fission; les bombes libérant cette énergie sont dites « nominales »), il faut une cinquantaine de générations, qui durent chacune en moyenne cinq milliardièmes de seconde, soit au total un quart de microseconde, puisque à chaque génération la puissance est multipliée par cinq ou six.

L'instant d'amorçage devra donc se situer environ une fraction de microseconde en amont de t_m, et l'on conçoit que la réalisation de cette performance présente quelques difficultés, ne serait-ce que parce que l'instant t_m n'est pas connu avec précision. N'insistons pas sur ce petit détail et remarquons que pour avoir des réactions se déroulant suivant le schéma prévu, il est nécessaire qu'il n'y ait pas de neutrons parasites qui fassent diverger le système avant l'instant d'amorçage théorique t_a.

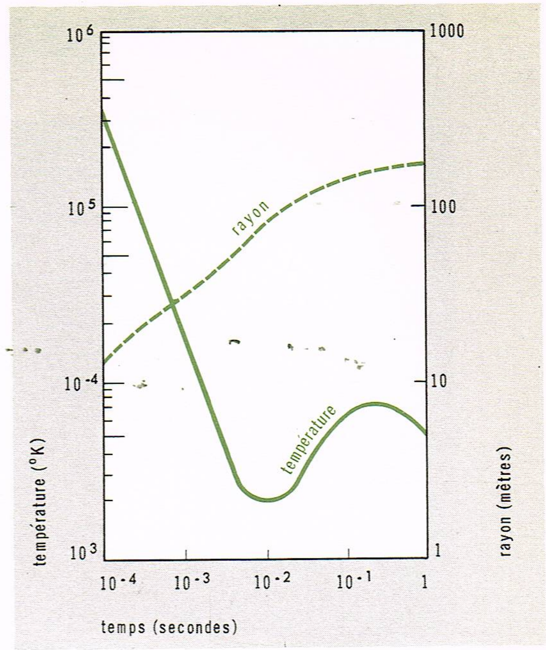
Actuellement, on ne dispose que de deux espèces de noyaux pouvant servir d'explosifs nucléaires : le plutonium 239 et l'uranium 235; il faut les utiliser presque purs : 98 p. 100 environ. Le plutonium est nucléairement plus intéressant que l'uranium 235. L'uranium 233, dont on ne dispose pas de quantités suffisantes, se situe entre les deux.

Le plutonium 239 est extrait par voie chimique des barreaux de combustible irradiés; ce plutonium comprend donc tous les isotopes qui résultent de la capture des neutrons par le Pu 239, entre autres le Pu 240. Or, ce dernier noyau présente un mode de désintégration spécial dit « fission spontanée »; sans cause apparente, les noyaux se fissionnent spontanément avec une période de cent milliards d'années. Malgré l'importance de ce chiffre, une molécule-gramme de Pu 240 émet un neutron toutes les deux microsecondes. Pour ne pas avoir de « mauvais amorçages », il faut réduire le nombre de neutrons dont l'émission n'est pas contrôlée, c'est-à-dire la masse de Pu 240, ainsi que le temps total pendant lequel il y a possibilité d'amorçage non commandé; il faut, au moins, que ce temps ne soit pas trop important par rapport au temps moyen qui sépare l'émission de deux neutrons parasites. Comme les masses critiques sont de l'ordre d'une dizaine de kilogrammes, on voit que si, dans une telle masse de 239, on a une molécule-gramme de 240, la proportion est de 2,5 p. 100. Des calculs simples montrent que, dans les barreaux de combustible, la proportion de Pu 240 dans le Pu 239 croît avec le temps; donc les barreaux qui ont été irradiés longtemps ont une teneur en 240 trop élevée et le plutonium extrait est impropre à la fabrication des bombes.

Finalement, nous voyons que ce phénomène de « fissions spontanées » fixe le

temps d'irradiation des barreaux dans les piles, puis une limite supérieure au temps de rapprochement des masses; ce temps doit être de quelques microsecondes, ce qui exige des pressions P énormes, que l'on obtient par la détonation d'explosifs conventionnels.

L'uranium 235, qui par ailleurs est nucléairement moins intéressant que le plutonium pour fabriquer des bombes, ne présente pas le phénomène de « fissions spontanées »; on peut donc concevoir des engins qui fonctionnent avec une faible charge d'explosif conventionnel, ce qui permet de réduire considérablement le poids et l'encombrement.



Température et rayon de la boule de feu, en fonction du temps après l'explosion.

une explosion, et ses effets

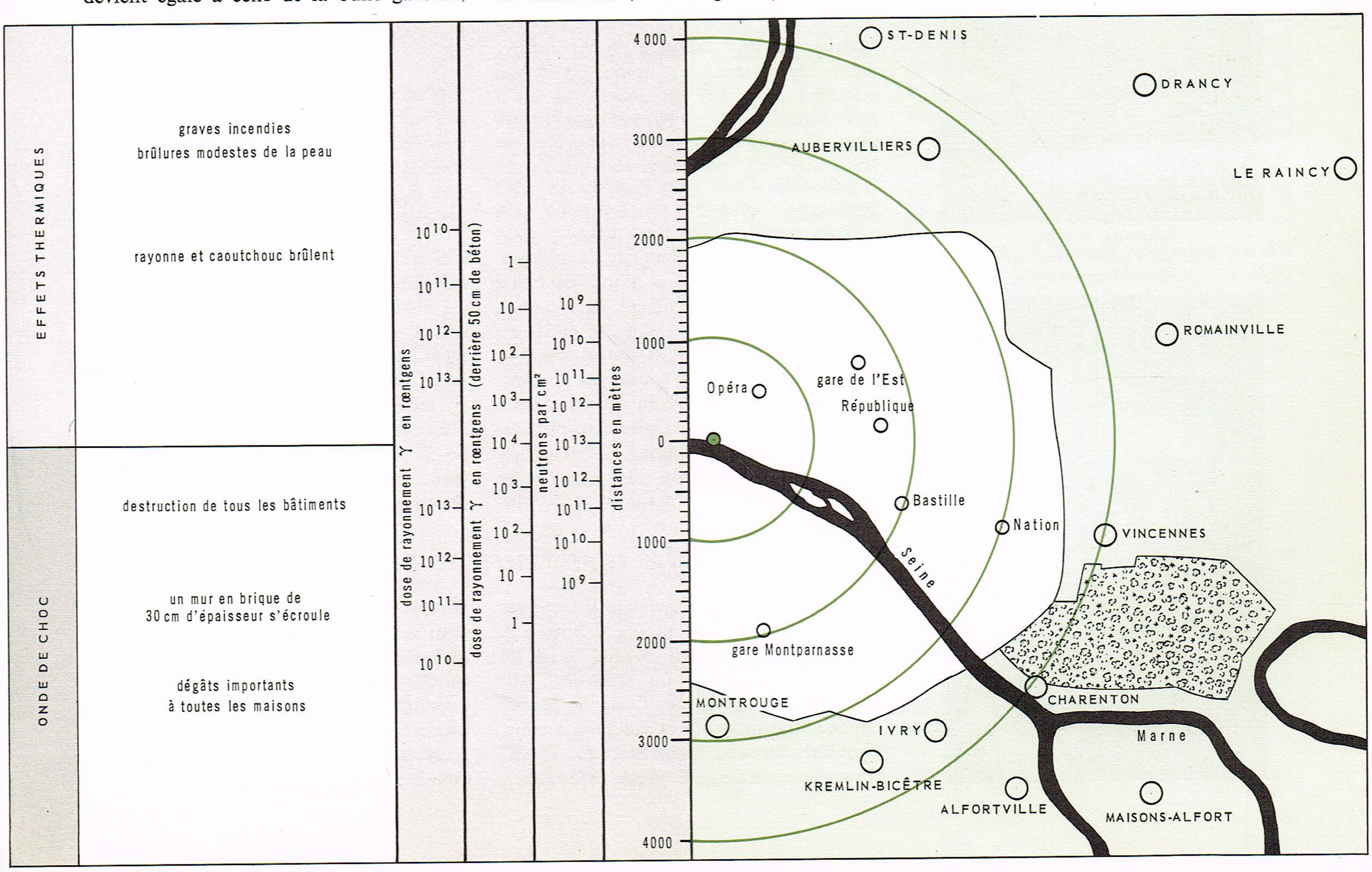
Immédiatement après l'explosion apparaît une sphère lumineuse dite « boule de feu », qui résulte de l'absorption par l'air des radiations électromagnétiques émises par les produits de fission, qui sont à des températures très élevées.

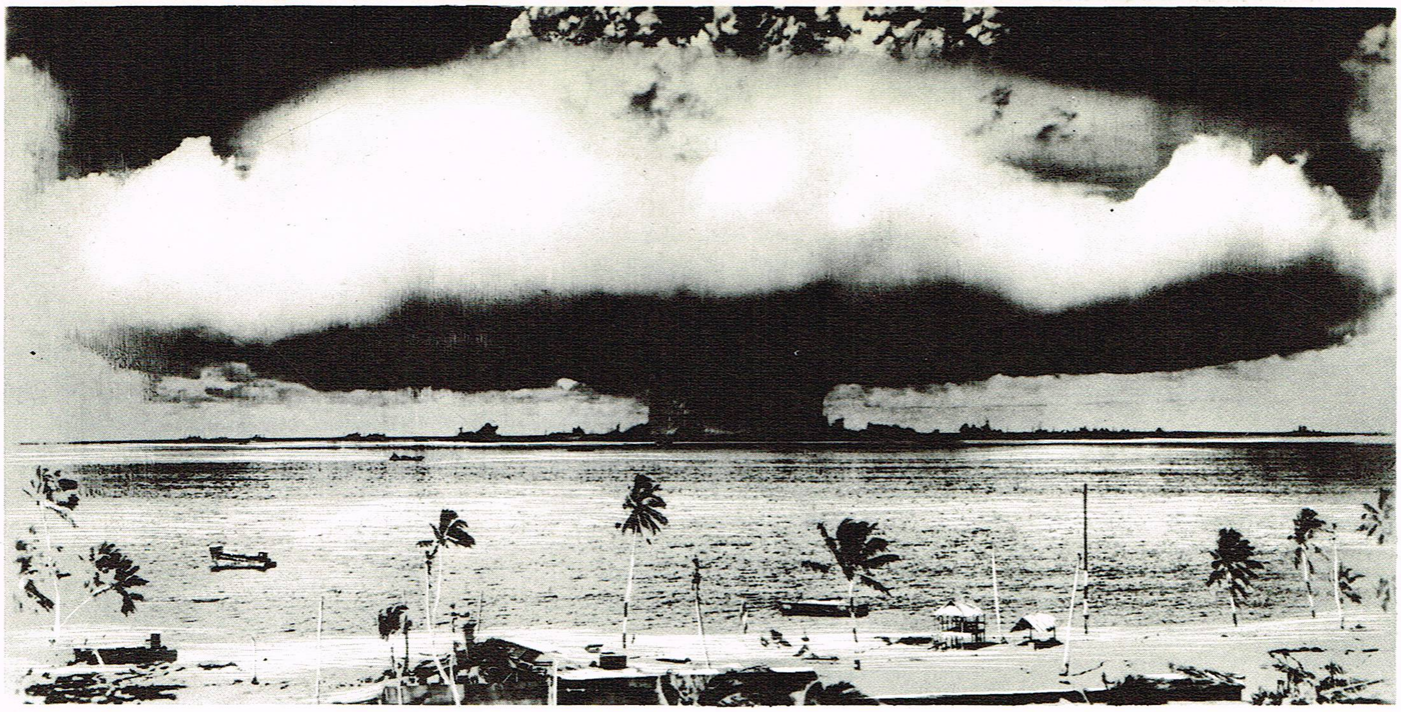
Cette boule de feu grossit et s'élève dans l'air, la température en est assez élevée pour vaporiser toutes les substances présentes, en particulier les produits de fission, l'uranium ou le plutonium non désintégrés, ainsi que l'enveloppe de la bombe. Quelque dix secondes après l'explosion, la boule n'est presque plus lumineuse. Tous les produits contenus à l'intérieur de la boule s'élèvent : au moment où la densité de l'air ambiant devient égale à celle de la bulle gazeuse, elle occupe, avec tous les produits radioactifs qu'elle contient, une étendue de plusieurs kilomètres : c'est le « champignon » qui restera visible jusqu'à ce que le vent l'ait dissipé.

Les particules radio-actives adhèrent aux poussières de l'atmosphère, en général plus grosses, et retombent progressivement sous l'influence de la pesanteur : c'est le phénomène dénommé « fall-out ».

Le principal effet destructeur d'une explosion nucléaire dans l'air est l'onde de choc. A Hiroshima, on estime qu'elle est responsable de 50 à 80 p. 100 des destructions (il y a eu des blessures sérieuses jusqu'à 3 500 m). Viennent ensuite les effets thermiques (20 à 30 p. 100), puis les radiations (15 à 20 p. 100).

Si, par exemple, une bombe atomique de 20 kt explosait à Paris au-dessus du Palais-Royal, le cercle passant par le Sacré-Cœur, la Bastille, l'Etoile et la place Denfert-Rochereau représenterait la superficie très endommagée par l'onde de choc; au-delà, il y aurait encore des incendies.





Explosion sous-marine d'une bombe A à Bikini (1946).



Brûlures causées par la bombe atomique d'Hiroshima (1945).



Les dégâts provoqués par une explosion sous-marine sont également considérables : tous les navires seraient vraisemblablement coulés dans un rayon de 800 m.

Les conséquences les plus importantes du rayonnement thermique sont l'inflammation de certaines matières combustibles qui, ultérieurement, provoque des incendies et les brûlures de la peau qui peuvent devenir mortelles dès qu'une certaine surface (environ 30 p. 100) est brûlée, même légèrement. Au Japon, on estime que 20 à 30 p. 100 des décès sont dus à cette cause (il y a eu des brûlures importantes jusqu'à 4 km).

L'explosion d'une bombe atomique s'accompagne d'émissions de radiations α, β, γ et de neutrons. Les neutrons sont émis uniquement au moment de l'explosion. Les produits de fission et la matière fissile qui n'a pas été fissée libèrent les α, β et γ. Une partie des γ est également émise au moment de l'explosion. Pratiquement, l'ensemble des radiations émises est conventionnellement subdivisé en deux parties : le rayonnement initial (première minute) et le rayonnement résiduel. Le rayonnement résiduel provient des produits de fission, des noyaux d'uranium et de plutonium non fissionnés et des corps divers qui ont été activés par les neutrons. Tous ces corps radio-actifs peuvent, au gré des vents, être transportés très loin du lieu de l'explosion.

Nous verrons que les lésions biologiques causées par les rayons γ sont consécutives à l'ionisation des molécules constituant les tissus. La dose létale qui entraîne la mort dans un délai de quelques jours est d'environ 400 röntgens.

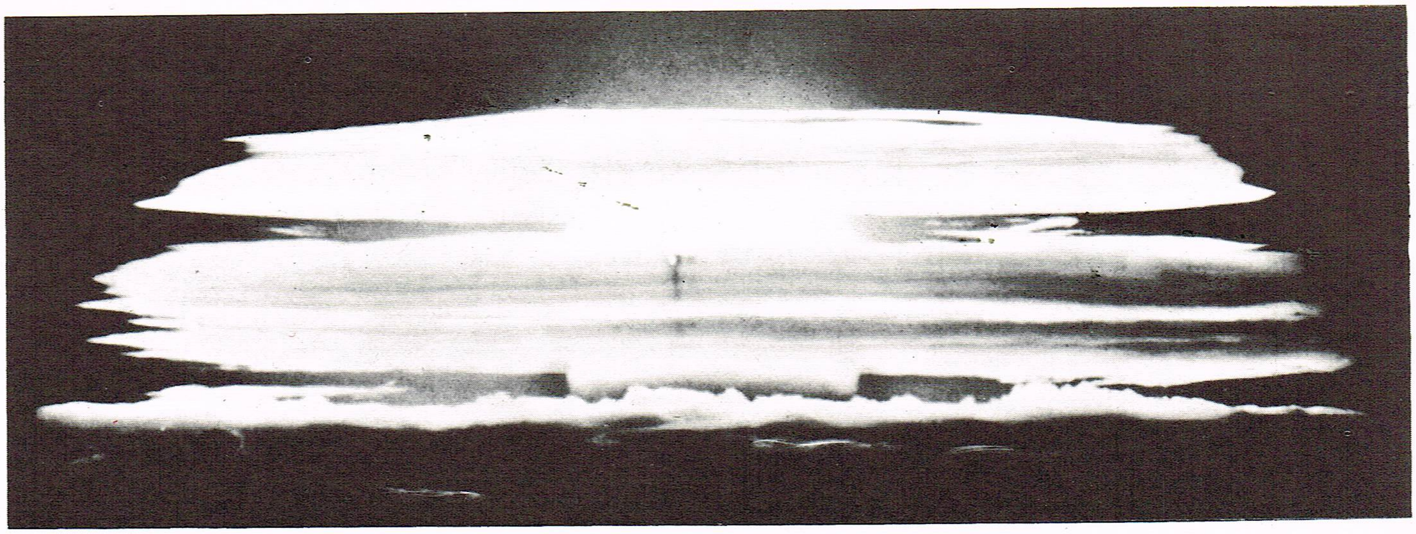
Les neutrons ne constituent pratiquement pas un danger supplémentaire, car les distances où leur intensité est mortelle sont inférieures à celles des γ. La protection la plus efficace pour les γ et pour les neutrons est le béton.

Les Etats-Unis réalisèrent la première explosion à Alamogordo, dans le désert du Nouveau-Mexique, le 16 juillet 1945; c'était une bombe au plutonium de 20 kt. Quelques jours plus tard (6 et 9 août 1945), deux bombes de 20 kt furent lancées sur le territoire japonais, à Hiroshima et Nagasaki, et hâtèrent certainement la fin de la Seconde Guerre mondiale. A ce jour, il y a eu environ 250 explosions (bombes A et H).

la terrifiante bombe H

Le 14 juillet 1949 les physiciens soviétiques faisaient exploser leur première bombe A et annonçaient ainsi au monde la fin du monopole américain. Le président Truman, le Pentagone, le Congrès délibérèrent longuement sur la conduite à tenir. L'amiral Strauss, président de l'AEC (Commission de l'Energie Atomique), adressa, le 5 octobre 1949, un mémorandum à ses collègues pour réclamer la réalisation d'une bombe thermonucléaire. Le 29 octobre 1949, sous la présidence de J. Robert Oppenheimer, un Comité d'experts se réunit qui rejeta à l'unanimité le programme proposé. Des raisons diverses furent exposées pour justifier ce refus :

Phot. U. S. I. S.



Explosion d'une bombe thermonucléaire à hydrogène à Bikini (21 mai 1956).

prix exorbitant, incertitude d'arriver au résultat escompté, et aussi crainte de voir se développer une redoutable course aux armements. Après de nombreuses discussions, le 31 janvier 1950, D. Acheson, secrétaire aux Affaires étrangères, et L. Johnson, secrétaire à la Défense, enlevaient la décision du président Truman. Après un travail intensif de trois ans, dirigé par Edward Teller, les recherches aboutirent à l'explosion de la première bombe H (Eniwetok, 31 octobre 1952) : c'était la première fois qu'une énergie d'origine thermonucléaire aussi massive (plus de 5 millions de tonnes de T. N. T.) était libérée par l'homme.

Les conditions nécessaires pour réaliser cette explosion thermonucléaire ne sont pas différentes de celles que nous avons déjà exposées : il faut chauffer, maintenir et confiner.

La température d'allumage était obtenue par l'explosion d'une bombe A, le confinement par inertie, la réaction choisie fut : D + T (deutérium et tritium). Bien que le tritium soit extrêmement cher et très radio-actif (période de douze ans), ce choix fut imposé par la petitesse de la température à atteindre : 50 millions de degrés seulement, au lieu de 400 millions pour la réaction D + D, ou 1 milliard pour $Li^6 + D$ (lithium et deutérium).

Les deux isotopes étaient utilisés sous forme liquide. A vrai dire, l'engin qui fonctionna n'avait rien d'opérationnel : c'était une machine très compliquée, qui pesait plus de 80 tonnes.

Pouvait-on simplifier et alléger? Les savants russes trouvent la solution et la cèdent involontairement; dix mois plus tard, le 12 août 1953, ils firent exploser leur première bombe H, qui était vraisemblablement opérationnelle et les Japonais trouvent du Li_6 dans les retombées radio-actives : immédiatement, on pense au deutérium de lithium DLi_6, qui est solide à la température ordinaire. Les réactions s'écrivent

$$Li^6 + n \rightarrow He + T$$
$$D + T \rightarrow He + n.$$

(Li, lithium; *n*, neutron; He, hélium; T, tritium; D, deutérium).

L'ensemble de ces deux réactions équivaut au

$$Li^6 + D \rightarrow 2He + 22{,}4 \text{ MeV}.$$

Cette réaction, qui exige un seuil du milliard de degrés, est en fait catalysée par le tritium, qui abaisse la température d'amorçage à 50 millions de degrés. Après cette découverte, et compte tenu de l'acquis américain, ce fut un jeu de construire une bombe opérationnelle; l'expérience eut lieu six mois plus tard, le 28 février 1954, à Bikini, et libéra 15 Mt, soit environ dix fois l'ensemble des bombardements alliés sur l'Allemagne au cours de la Seconde Guerre mondiale.

C'est à la suite de cette expérience que des poussières de l'atoll de Bikini, qui avaient été activées par l'explosion, furent transportées par les vents à plus de 300 km et irradièrent les malheureux marins du bateau de pêche japonais *Fukuryu-Maru*.

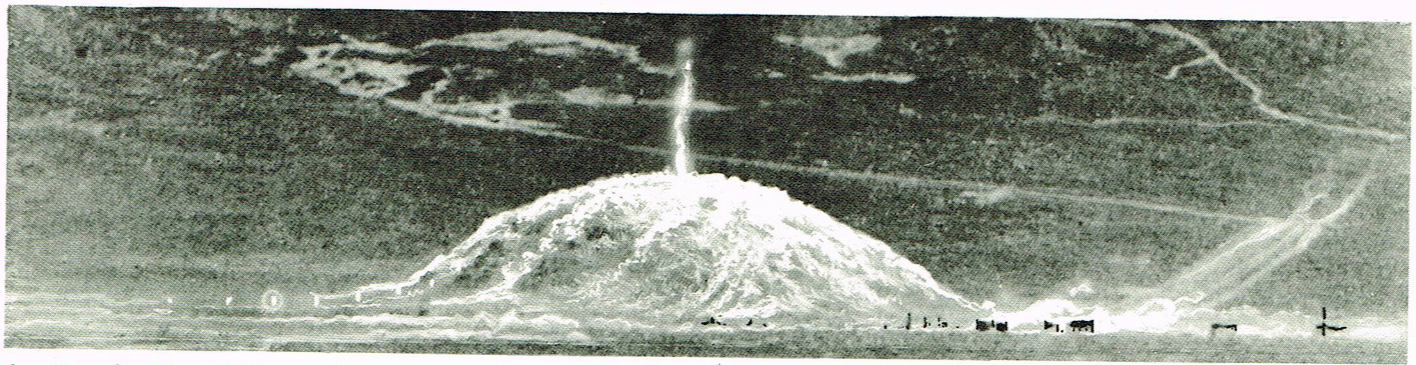
le sommet du « progrès » : les bombes 3 F

Ce premier type d'engin au deutériure de lithium fut perfectionné par les Russes, puis par les Américains, qui aboutirent tous deux aux bombes dites 3 F (Fission-Fusion-Fission) : une bombe A sert d'amorce, fonctionne par fission, puis allume des réactions thermonucléaires avec D Li^6; certains des neutrons créés provoquent des fissions dans une « couverture » d'U 238 qui entoure le tout et qui sert en même temps à confiner. Les bombes 3 F sont « plus sales », c'est-à-dire libèrent plus de produits radio-actifs que les bombes 2 F.

Comme la bombe H est déclenchée par une bombe A, les dégâts de toute nature (onde de choc, onde thermique ou radiations) sont beaucoup plus importants que ceux qui sont produits par une simple bombe A.

L'effet de souffle augmente comme la racine cubique de la puissance. Une bombe H de 20 Mt comme celles que semblent transporter les avions du Strategic Air Command a mille fois la « puissance nominale »; elle pourrait donc détruire les bâtiments dans un rayon de 15 km. Les plus grandes cités du monde pourraient être complètement anéanties par une seule bombe H.

L'effet de température augmente comme



Au Nevada (Etats-Unis), une explosion nucléaire souterraine a fait surgir cette véritable colline qui apparaît au premier plan de la photographie.

la racine carrée de la puissance. La surface soumise à ces dégâts est donc proportionnelle à la puissance; on peut donc admettre que cette même bombe H de 20 Mt produirait, par élévation de température, des dégâts extrêmement importants dans un rayon de 30 km dans le cas où l'explosion a lieu près du sol; mais les spécialistes ont calculé que, pour utiliser rationnellement de telles bombes, il faudrait les faire exploser à 25 ou 30 km; un vaste incendie embraserait alors une étendue de terrain circonscrite par un cercle de 100 à 200 km de rayon. On retrouve ainsi le chiffre souvent cité, à savoir qu'une douzaine de bombes, judicieusement réparties, seraient suffisantes pour anéantir complètement la France.

Devant ces chiffres, le danger des radiations nucléaires émises au moment de l'explosion fait figure de quantité négligeable : toutes les personnes qui se trouvent dans une situation où elles risquent de recevoir la dose létale auraient « déjà » été tuées par l'onde de choc ou par les effets thermiques. Seul le rayonnement résiduel pourra continuer à exercer ses ravages longtemps après l'explosion, et souvent loin du point où elle s'est produite.

La bombe H a d'autant plus de succès que certaines personnalités ont montré qu'elle était, à puissance égale, bien meilleur marché que les bombes conventionnelles ou les bombes A.

Des experts américains estiment que la mégatonne produite par un tel engin avait un prix de revient qui était inférieur au millième de celui des explosifs conventionnels.

des bombes susceptibles d'applications pacifiques ?

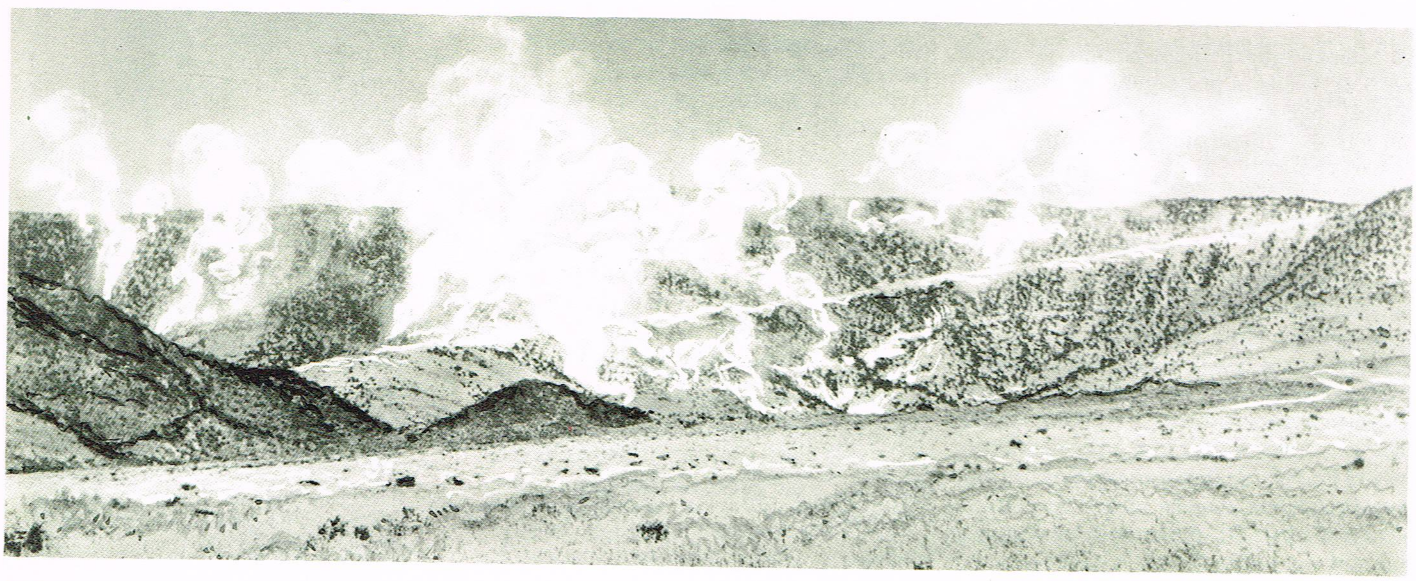
Mais à côté du cynisme de ces chiffres et de ces visions apocalyptiques, un espoir surgit, car on a découvert de nombreuses applications pacifiques de l'énorme énergie libérée par les réactions de fusion.

On pourrait l'utiliser pour modifier les conditions climatiques de certaines régions, pour creuser des canaux, des lacs ou des barrages artificiels...

Très récemment, Edward Teller, père de la bombe H américaine, a dit : « L'emploi de ces bombes pour l'exécution des grands travaux pourrait devenir le premier mode d'utilisation massif et réellement économique de l'énergie atomique. »

Mais de toutes ces conceptions issues d'imaginations fertiles, celle qui semble la plus riche en promesses est la chaudière souterraine ouverte dans les profondeurs de la Terre : un grand nombre de problèmes qui hantent les ingénieurs chargés d'étudier et de réaliser des centrales nucléaires seraient alors résolus.

Nuage de poussières non radio-actives provoqué par une explosion nucléaire souterraine, au Nevada (Etats-Unis), en 1957.



Phot. U. S. I. S.

méfaits et bienfaits des rayonnements nucléaires

Qu'entend-on par « rayonnements nucléaires » ? Nous admettons, comme première définition, que c'est l'ensemble des « ondes » et « particules » légères libérées au cours des réactions nucléaires. Ces rayonnements, nous l'avons vu, appartiennent principalement à quatre groupes : α, β, γ et neutrons. Les radiations α et β sont absorbées par de faibles épaisseurs de matière, mais il n'en est pas de même pour les γ et les neutrons.

de mortels effets si l'on n'y prend garde...

Quel est le mécanisme général d'action de ces particules ? Quand elles se déplacent dans la matière inerte ou vivante, la plus grande partie de leur énergie est utilisée pour ioniser les atomes ou les molécules qu'elles rencontrent, c'est-à-dire pour leur arracher un ou plusieurs électrons. Comme la perte d'un électron confère à ces atomes une puissante aptitude à réagir chimiquement avec les acides nucléiques des cellules vivantes, on comprend que, si certaines radiations produisent des taux d'ionisation très différents de ceux qui sont nécessaires pour assurer le développement normal des cellules, on arrive à créer des perturbations qui, au-delà d'une certaine importance, entraînent la mort de l'individu; si ces perturbations affectent les gènes, il pourra y avoir création d'une espèce différente : c'est ce qu'on appelle une « mutation ».

Si tous les savants sont à peu près d'accord sur les phénomènes de base qui expliquent l'action des rayonnements sur les êtres vivants, il n'y a plus d'unanimité quand il faut en chiffrer les conséquences. Il semble néanmoins :

1° que les doses « sûrement mortelles » augmentent au fur et à mesure que l'on descend dans le degré de l'organisation des êtres vivants. Tandis que 600 röntgens administrés en une seule fois sont, en général, suffisants pour provoquer la mort d'un homme, il en faut cent fois plus pour une mouche, et des centaines de milliers de röntgens sont nécessaires pour une cellule isolée, et encore beaucoup plus pour détruire certains éléments de la cellule. Il semble donc que la vie est un état de la matière d'autant plus fragile qu'il est plus organisé;

2° que la dose mortelle pour l'homme correspond à environ 60 000 ergs, soit la cent-millième partie de ce que notre organisme consomme d'énergie en une seconde;

3° que, dans un grand nombre de cas, le phénomène présente un seuil : une dose reçue pendant quelques secondes peut être très dangereuse, alors que la même dose accumulée au cours d'un temps beaucoup plus long par doses unitaires faibles séparées par un certain temps de relaxation, ne produit pas d'effets sensibles. La matière vivante est caractérisée par de grandes facultés d'adaptation et de récupération, de sorte que les effets d'agressions séparées par des périodes de repos ne se cumulent pas ;

4° que sur des groupes bien déterminés d'êtres vivants, le phénomène est subjectif, c'est-à-dire qu'il peut varier sensiblement d'un individu à un autre ;

5° que les gènes qui constituent les chromosomes des cellules reproductrices sont atteints les premiers, et un individu peut, après irradiation, sembler tout à fait normal et avoir une descendance tarée. L'action des radiations sur les gènes produit souvent une mutation biologique.

En fait, on a peu de résultats certains quant à l'action des rayonnements sur les organismes humains, et il y a un énorme coefficient d'incertitude quand on veut étendre à l'homme les résultats trouvés sur des animaux. Mais il ne faudrait sous aucun prétexte que cette incertitude conduisît à minimiser le danger et à transgresser les règles de sécurité.

Mais comment ces rayonnements nucléaires peuvent-ils atteindre un organisme vivant ? Pratiquement, il faut distinguer deux types principaux d'irradiations, suivant que la source de rayonnements se trouve à l'extérieur ou à l'intérieur du corps vivant (c'est uniquement dans le second cas que l'on parle de contamination). L'irradiation externe se produit quand la source est éloignée de l'organisme; elle cesse dès que l'on interpose une protection ou que l'on s'éloigne suffisamment de la source. L'irradiation interne provient de produits radio-actifs qui se trouvent à l'intérieur du corps, où ils ont pénétré par inhalation d'air contaminé ou par des lésions cutanées. Les dangers de l'irradiation interne proviennent du fait qu'elle est permanente; elle ne diminue qu'avec la décroissance radio-active de l'émetteur et l'élimination naturelle consécutive à la vie de l'organisme. Il y a finalement deux « périodes » qui jouent : la période de l'élément radio-actif (c'est-à-dire la durée au bout de laquelle la moitié des atomes auront subi une transmutation) et celle de l'élimination biologique (durée au bout de laquelle la moitié des corps absorbés auront été éliminés d'une manière ou de l'autre). Il est évident que ce sont les éléments à vie intermédiaire et susceptibles d'être retenus longtemps par les organes qui seront les plus dangereux.

On peut maintenant se demander comment se créent ces émetteurs de rayonnements nucléaires, qui peuvent créer des troubles si graves en nous irradiant.

Il y a d'abord des corps radio-actifs naturels qui se trouvent à la surface du globe et qui créent en permanence un « bain de rayonnements » auxquels nous ne pouvons échapper. On admet généralement qu'un homme au cours de sa vie normale reçoit ainsi des doses qui, accumulées, représentent environ 40 röntgens, mais qui risquent de s'accroître si l'on augmente les examens radiologiques. Actuellement, à cette radio-activité naturelle s'ajoute la radio-activité artificielle résultant tant des explosions nucléaires que de l'exploitation des réacteurs nucléaires. On a trouvé que la moyenne de radio-activité

dans l'air, au niveau du sol, attribuable aux explosions de bombes de tous types, ne représentait que quelques pour cent de la moyenne naturelle. Mais il y a de très fortes variations autour de cette moyenne; il est bien évident que la proximité des champs de tir où ont lieu les explosions atomiques peut être dangereuse; et si nous pensons aux effets radiologiques d'une guerre thermonucléaire à grande échelle, le tableau est très différent et offre des possibilités de visions apocalyptiques.

...et quelques bienfaits

Pensons de nouveau à ces corps radioactifs artificiels dont les noyaux se disloquent périodiquement en émettant des particules qui peuvent être détectées et comptées, qui peuvent plus ou moins facilement traverser la matière et qui ont également la possibilité de détruire ou de modifier les structures vivantes. Voilà beaucoup de possibilités offertes par ces particules infiniment petites, et en donnant libre cours à notre imagination on peut imaginer une infinité d'applications.

Le pouvoir destructeur peut, dans certains cas, être recherché; en agronomie ou dans l'industrie de la conserve, de très nombreuses études ont été effectuées pour utiliser les radio-éléments afin de stériliser par la destruction de certains micro-organismes; en thérapeutique, sous réserve de contrôler très soigneusement leur utilisation, on peut s'en servir pour détruire des cellules malignes.

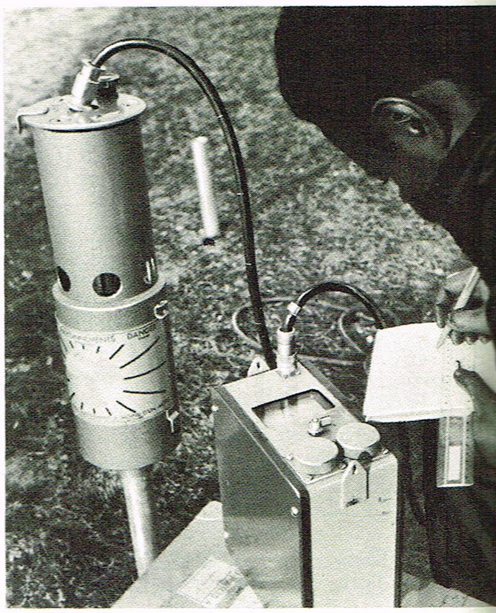
La possibilité de produire des mutations est aussi utilisée en agronomie pour créer de nouvelles espèces présentant des avantages certains sur les anciennes — nouvelles qualités de riz en Camargue, par exemple. Dans le même ordre d'idées, on a souvent constaté que certaines réactions chimiques étaient très nettement accélérées par des rayonnements nucléaires, qui jouent ainsi un rôle de catalyseur.

Les caractéristiques de ces particules font qu'il est possible de concevoir des appareils de mesure ayant des possiblités exceptionnelles : mesurer des épaisseurs de tôle et commander automatiquement des rouleaux de laminoir, mesurer des niveaux de liquide dans de grands réservoirs, mesurer des taux d'usure de segments de moteur ou de revêtements routiers, mesurer des débits...

La haute perfection aujourd'hui atteinte par les appareils qui détectent et comptent ces particules en fait des auxiliaires souvent indispensables. Grâce à eux, on peut suivre à la trace le développement de certains éléments. Ce principe, appliqué à la biologie et à la physiologie, a permis de très grands progrès dans la connaissance du corps humain, de sa composition, des fonctions de certains organes; détection, localisation et diagnostic sont grandement facilités. En agronomie, des mesures jadis impensables ont pu être effectuées : rythme de la croissance des plantes, montée de la sève, échanges des engrais avec le sol...

Les molécules marquées, en général molécules de chimie organique marquées au C^{14}, ont permis d'étudier avec efficacité la structure et la synthèse de corps organiques particulièrement complexes.

On est obligé d'arrêter l'énumération. Disons simplement qu'une étude faite en 1958 a montré qu'en France les radio-isotopes avaient déjà permis de réaliser des économies se chiffrant à près d'un milliard de francs.





Rayonnements émis.

Une masse donnée d'un corps radio-actif émet des particules. Comptons toutes ces particules émises dans toutes les directions pendant une seconde; leur nombre N caractérise l'intensité du rayonnement. Comme ce nombre est en général très grand, on l'exprime en « curie », unité arbitraire qui vaut 37 milliards de particules par seconde. Dire qu'une masse d'un corps a une activité d'un millicurie signifie que cette masse émet 37 millions de particules par seconde. Un gramme de radium a une activité très voisine d'un « curie ».

Rayonnements reçus, autres que X ou γ.

Considérons un émetteur γ et un cube d'air d'un centimètre de côté, contenant 0,001 293 g d'air.
Les rayons γ qui traversent le cube provoquent des ionisations, c'est-à-dire créent des charges électriques. Au bout d'un certain temps, la somme des charges créées est égale à une unité C. G. S.; par définition, la masse d'air contenue dans le cube aura absorbé une dose égale à un röntgen. La création de cette quantité d'électricité par ionisation exige une énergie de 84 ergs par gramme d'air.
L'expérience montre que si, avec une source déterminée, on irradie pendant un temps déterminé la même masse d'air, d'eau ou de matière vivante placée dans la même position par rapport à la source, l'énergie absorbée est approximativement la même dans les trois cas.

Rayonnements reçus, cas des rayons X ou γ.

Dans ce cas, la quantité de rayonnement absorbé s'exprime en « rad », qui équivaut à 100 ergs absorbés par gramme de matière.

Photothèque, phot. C. E. A.

Mesure de l'humidité du sol à l'aide d'une sonde à neutrons.

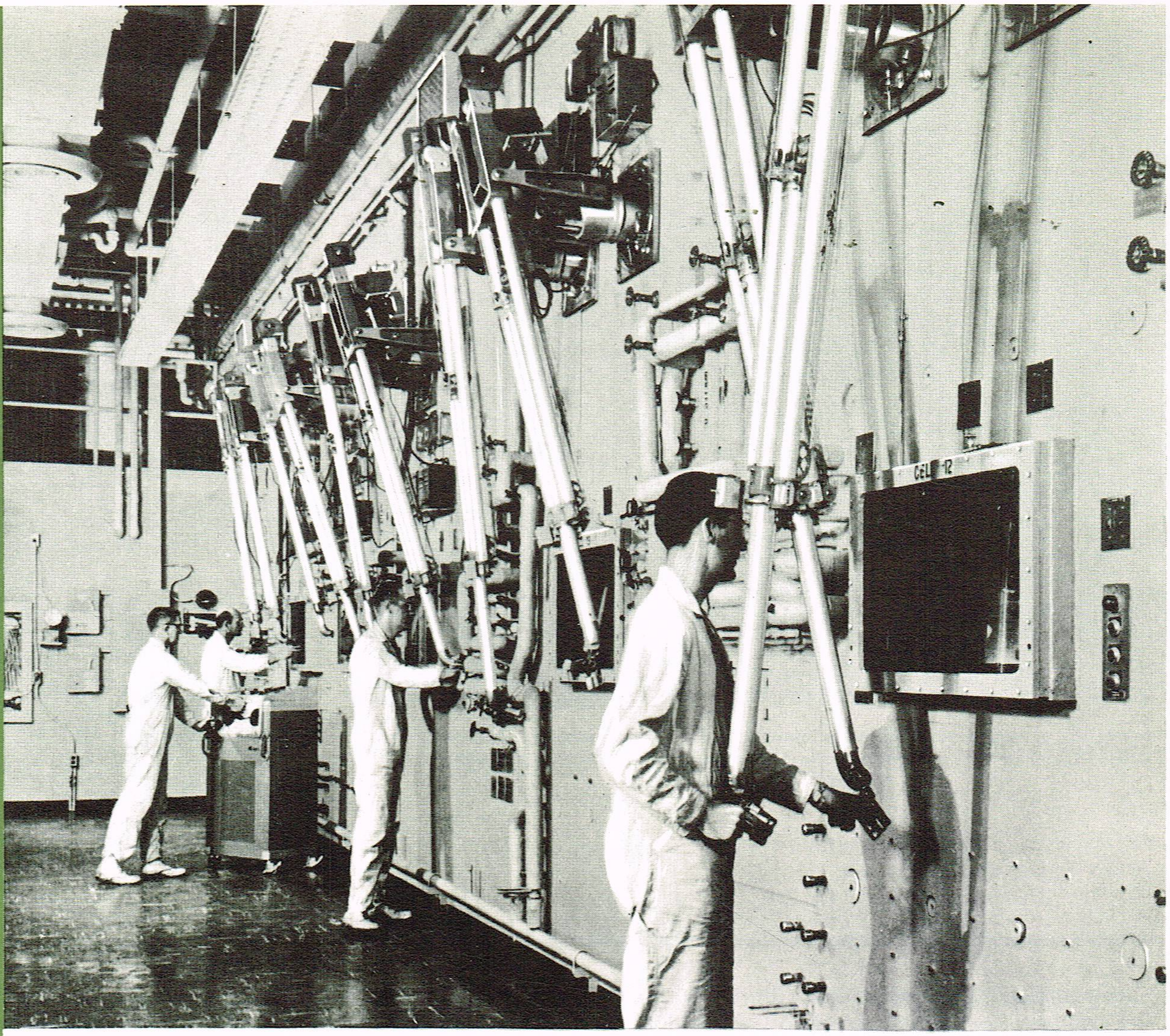
◀

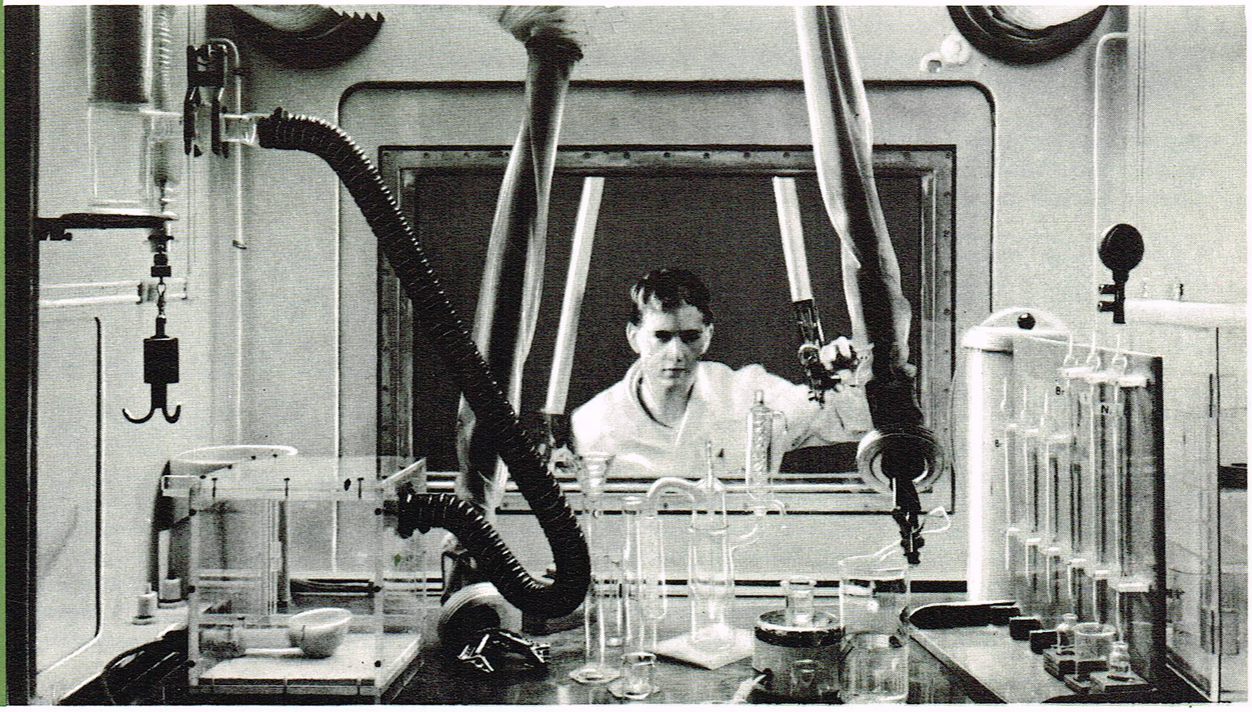
Recherche et localisation de fuites de gaz sur un réseau de distribution du Gaz de France.

◀

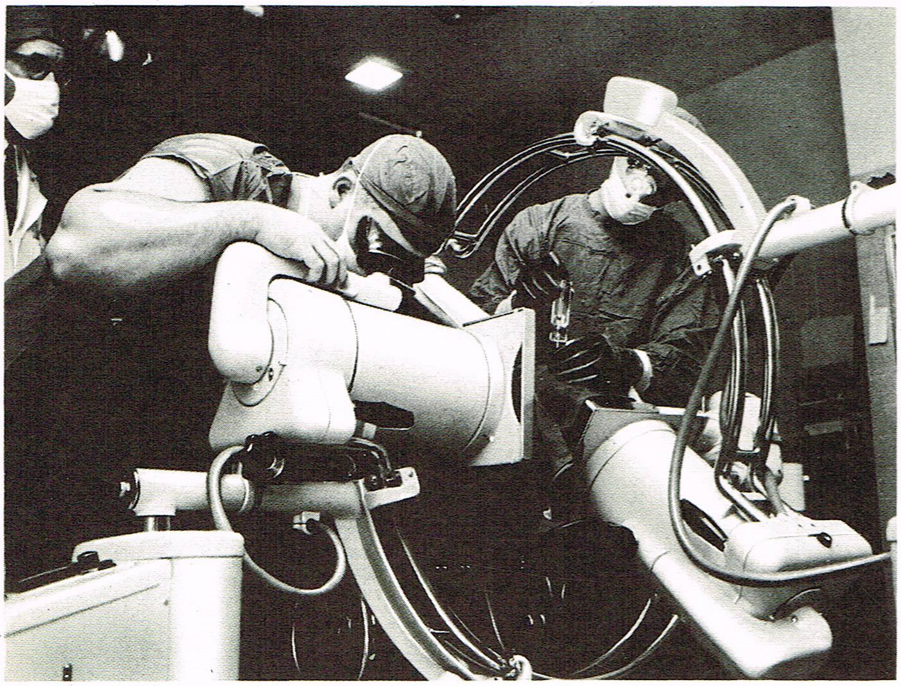
Préparation d'isotopes radio-actifs au moyen de télécommandes.

▶

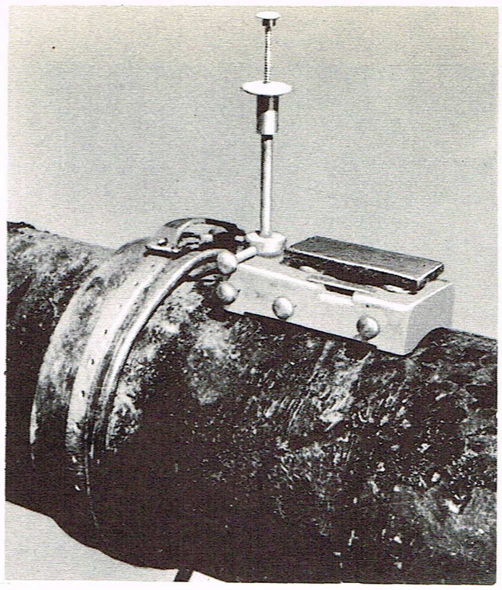




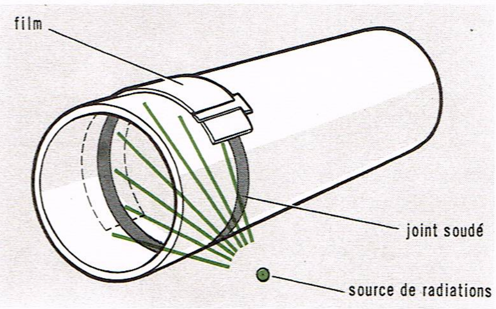
Phot. U. S. I. S.



Implantation d'yttrium 90 dans la glande pituitaire d'un patient atteint d'une tumeur maligne.



Contrôle non destructif d'une soudure au moyen d'une source de rayons γ.



Il y a bien longtemps de cela, le roi Agamemnon est assassiné dans son palais à son retour de la guerre de Troie, les Corinthiens envahissent Argos. Au milieu des cris d'horreur, la servante Narsès demande comment s'appelle le spectacle qu'elle voit. Un mendiant tourné vers l'avenir lui répond : « Cela a un très beau nom, femme Narsès. Cela s'appelle l'Aurore. »

La même réponse ne pouvait-elle pas être donnée lorsque les habitants d'Hiroshima virent tôt dans la matinée du 6 août 1945 une lumière aveuglante illuminer leur ciel?

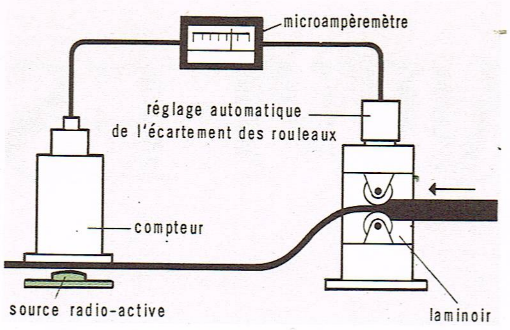
Rapidement, le monde entier fut convaincu que la nouvelle source d'énergie qui était désormais à la disposition des hommes allait modifier leur façon de vivre et de penser. Il est maintenant possible de faire un premier bilan de ce « quelque chose » d'inédit que ce jour apporta à l'histoire.

Ce fut d'abord l'obligation, que perçurent plus ou moins rapidement et plus ou moins intensément les responsables, de renoncer — sous peine de destruction presque complète de la civilisation humaine — à certains types de « solution » pour résoudre les conflits qu'entraîne nécessairement l'interaction des libertés et également d'essayer d'instaurer une communauté mondiale afin d'assurer la conservation de ce bien très précieux : la paix, du moins si l'on entend par ce dernier mot non la fin de tout conflit, mais le refus de résoudre les divergences d'intérêts ou d'idéals par des moyens militaires.

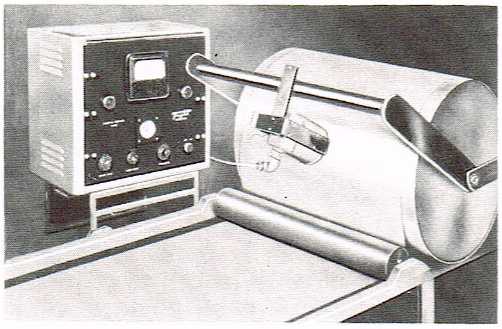
Il importe de noter le rôle qu'ont joué les scientifiques de tous les grands pays dans l'élaboration d'une mentalité rendant possibles plusieurs moratoires atomiques, et il faut espérer qu'ils deviendront de plus en plus ambitieux.

Bien que des impératifs militaires aient permis que la science nucléaire ait un développement inespéré, il ne faut pas oublier que « réaction nucléaire » ne signifie pas uniquement « bombe ». Ces réactions mettent également à la disposition de l'humanité une réserve d'énergie extrêmement importante. Et nous savons toute l'importance que prend un tel fait à l'heure même où les ressources énergétiques conventionnelles commencent à s'épuiser. Il semble à peu près évident que d'ici une quinzaine d'années, quand on aura résolu tous les problèmes qui restent encore en suspens, les centrales nucléaires seront appelées à un grand développement, car elles utilisent comme « combustible » les atomes d'uranium, qui ne peuvent avoir

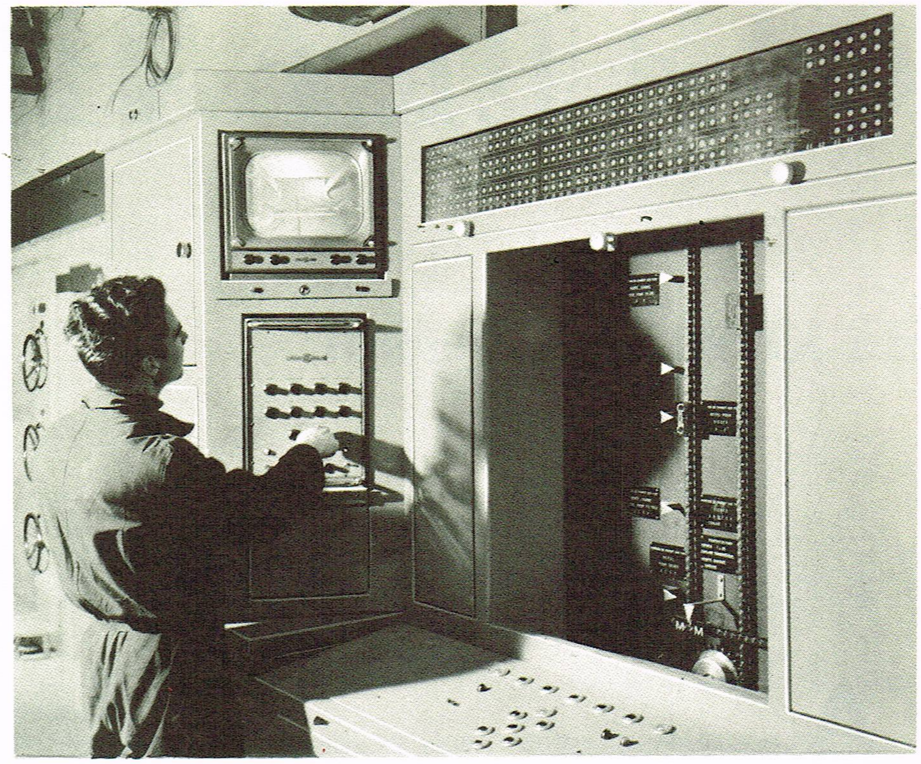
Phot. U. S. I. S. Doc. « Esso ».



Contrôle automatique de l'épaisseur des tôles.



Contrôle de l'épaisseur d'une feuille de caoutchouc au moyen des rayons β (appareil Tracerlab).



A Marcoule, la télévision en circuit fermé permet de contrôler le processus d'extraction automatique du plutonium contenu dans les barreaux d'uranium des réacteurs nucléaires.

d'autre utilité; par contre, on a le plus grand intérêt à économiser les combustibles fossiles, qui peuvent servir à beaucoup d'autres usages, par exemple à fabriquer des matières plastiques.

Il n'est pas nécessaire d'insister sur ce que la science nucléaire a apporté à la compréhension des lois qui gouvernent la nature et l'organisation du monde. Quel coup de fouet a connu, grâce à elle, la recherche scientifique dans tous les domaines! Elle a fait progresser d'une manière spectaculaire un grand nombre de techniques : métallurgie, électronique, biochimie...

Un climat de recherche a ainsi été créé, les gouvernements ont reconnu la nécessité de consacrer à la recherche des crédits substantiels; ils ont pris conscience que les savants représentaient pour la nation un bien précieux; peu à peu, le capital « homme » prend le pas sur le capital « argent ».

La complexité des études à conduire, la qualité des cerveaux nécessaires, l'importance des installations, les risques de toute nature que l'on court ont immédiatement montré que tout cela était hors des possibilités des sociétés privées, même les plus puissantes, et que seuls les Etats importants pouvaient s'attaquer à de tels problèmes. Et, rapidement, il a fallu admettre que même les grands Etats n'étaient pas de taille à supporter tous les efforts et qu'il fallait former des associations internationales, se réunir fréquemment pour faire le point, confronter les points de vue et échanger des idées.

L'atome, qui pourrait détruire les hommes si l'énergie qui est accumulée dans son sein était libérée inconsidérément, peut ainsi créer un lien qui va unir ces mêmes hommes pour faire progresser leurs connaissances.

Le développement pacifique de l'énergie atomique est la plus grande aventure de l'humanité; maintenant elle est vécue par le monde entier, qui y est engagé d'une manière irréversible, même si, comme toutes les aventures, elle comporte nécessairement des risques.

La révolution nucléaire apparaît à un moment de l'histoire où les conditions économiques et politiques ne sont pas du tout comparables à celles d'autrefois; les répercussions de l'utilisation de l'atome risquent donc d'être d'autant plus profondes et peut-être plus violentes. L'homme aura besoin de faire preuve de beaucoup de sagesse pour les contrôler.

Photothèque.

LES HAUTES

Au cours des premiers mois de l'année 1963, la presse spécialisée et la grande presse nous révélaient un exploit réalisé par un groupe de chercheurs soviétiques : ces savants avaient réussi à chauffer, pendant quelques centièmes de seconde, un volume gazeux de quelques millimètres cubes à une température de 40 millions de degrés. Un double record était ainsi battu : celui de la température atteinte et celui de la durée de maintien de cette température.

Performance étonnante permettant de reproduire au laboratoire les températures régnant dans l'atmosphère des étoiles. Mais performance bien modeste eu égard au but poursuivi, la fusion thermonucléaire contrôlée mettant en jeu les températures cent fois plus élevées réalisées au cœur de ces mêmes étoiles.

De sorte que cet exploit n'est qu'une nouvelle étape de la course vers les plus hautes températures, course commencée, au dire des spécialistes, depuis près de 800 000 ans : le sinanthrope savait faire du feu et l'homme de Neandertal l'utilisait déjà avec compétence.

A partir de cette époque, les différentes phases du développement de l'humanité correspondent sensiblement aux températures maximales que l'homme a su réaliser. La première expérience « haute température » de l'homme se traduisit par la cuisson des aliments, destinée d'abord à rendre leur digestion plus facile, puis à les conserver. Mais sa deuxième expérience, toujours aux temps néolithiques, permit la fabrication de poteries pour le stockage des aliments et des graines.

La conservation des poteries jusqu'à l'époque actuelle indique bien que ces récipients avaient subi non un simple séchage au soleil, mais bien une cuisson jusqu'à des températures de l'ordre de 500 °C ou même de 900 °C. La cuisson avait été effectuée non plus sur un feu de camp, mais dans un véritable four. Cette première transformation chimique, celle de l'argile, fut suivie de bien d'autres à mesure que le seuil des températures atteintes augmentait. L'âge du bronze, dont l'élaboration nécessitait des températures à peine supérieures à celles de la céramique, dura deux millénaires. L'âge du fer, qui lui succéda et dure encore, commença il y a trente siècles, lorsqu'on parvint à obtenir des températures bien supérieures à 1 000 °C.

La métallurgie ouvrait l'ère de la civilisation industrielle. Celle-ci a débuté en fait au moment où l'homme a su employer les métaux et les combustibles à haut pouvoir calorifique, pour utiliser la chaleur autrement que comme agent de chauffage. Plus tard, la transformation de la chaleur en d'autres formes d'énergie a été un facteur fondamental pour l'évolution de l'humanité.

TEMPÉRATURES

par
Robert COLLONGUES

du feu à la bombe H

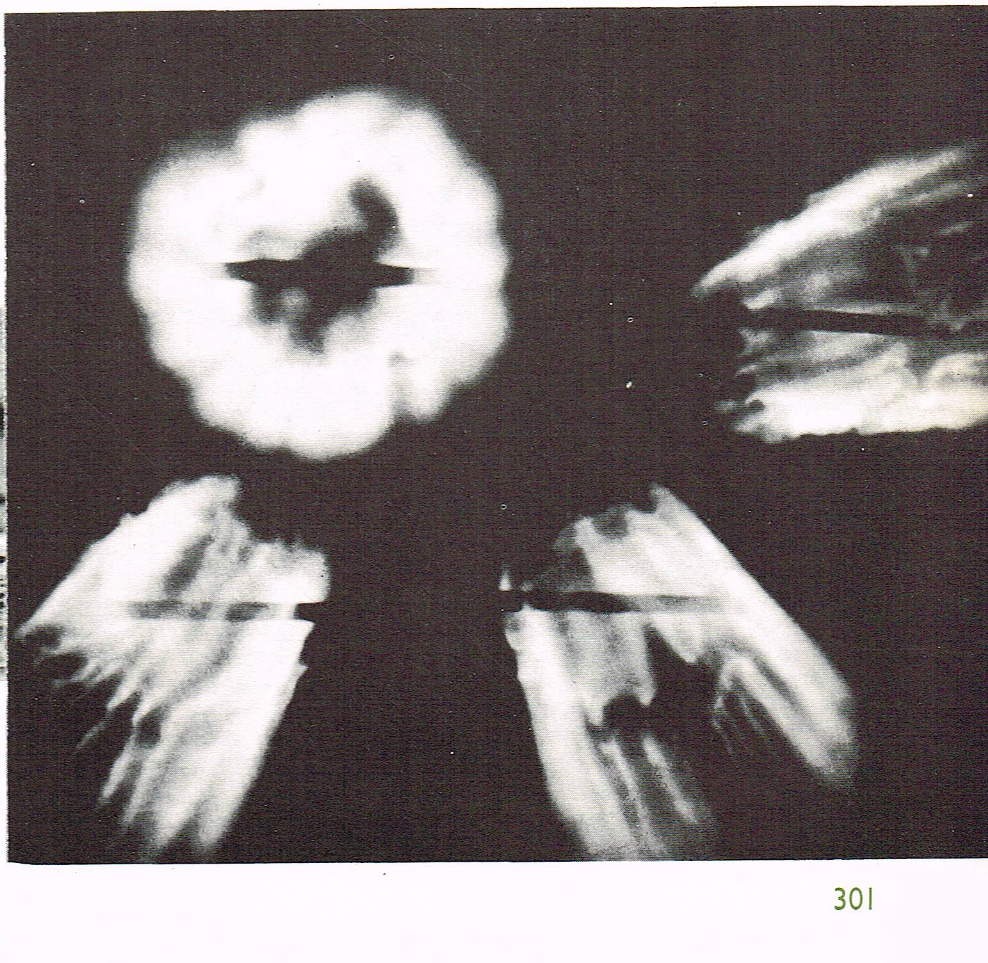
Les principes de la thermodynamique nous permettent de préciser la notion de chaleur à l'aide de trois propositions.

La chaleur est une forme particulière d'énergie. Or, dans l'expression de toute forme d'énergie intervient un produit de deux facteurs : un facteur de qualité et un facteur de quantité. L'énergie fournie par une chute d'eau est le produit de la hauteur de la chute (qualité) par son débit (quantité). L'énergie mécanique est le produit de la pression par la variation de volume. L'énergie électrique est le produit d'une variable de qualité, la tension, par une variable de quantité, la quantité d'électricité.

De la même manière, l'énergie calorifique sera un produit de deux facteurs. Le facteur de qualité sera évidemment la température. La variable de quantité a reçu le nom d'« entropie». Essayons de la définir à l'aide de la deuxième proposition.

La chaleur est une forme désordonnée de l'énergie. Considérons par exemple une balle de fusil en mouvement.





Des centaines de milliers d'années séparent l'apparition des deux techniques illustrées par ces photographies : l'obtention du feu de bois par frottement et le confinement dans une « bouteille magnétique » de plasma se trouvant à des millions de degrés.

Phot. Crevaux, G. A. E. (Zurich).

Au cours de son déplacement rapide et orienté, elle transporte une certaine forme d'énergie : l'énergie cinétique. La quantité d'énergie transportée est proportionnelle à la masse du projectile et au carré de sa vitesse; son expression s'écrit $E = \frac{1}{2} m V^2$.

Si cette balle heurte un obstacle, par exemple une plaque de blindage, son énergie cinétique se transforme en chaleur, la quantité de chaleur dégagée pouvant être telle que l'on observe une fusion partielle de la balle ou de la plaque. Au cours de cette transformation, les atomes abandonnent leur mouvement ordonné dans la balle pour se mettre à bouger d'une façon totalement désordonnée à l'intérieur de l'ensemble balle-plaque. La transformation d'une énergie ordonnée en une énergie désordonnée provoque un dégagement de chaleur. Mieux encore, considérons un gaz enfermé dans un corps de pompe. La théorie cinétique nous apprend que les atomes de gaz sont animés de mouvements incessants. C'est un système désordonné. Fournissons-lui de l'énergie, par compression par exemple. On sait que l'on communique ainsi de la chaleur. Dans ce cas la chaleur apparaît en fournissant de l'énergie à un système préalablement désordonné.

La notion de désordre est donc inséparable de la notion de chaleur. L'entropie, variable d'intensité dans l'expression de la chaleur, exprimera précisément la quantité de désordre contenue dans le système.

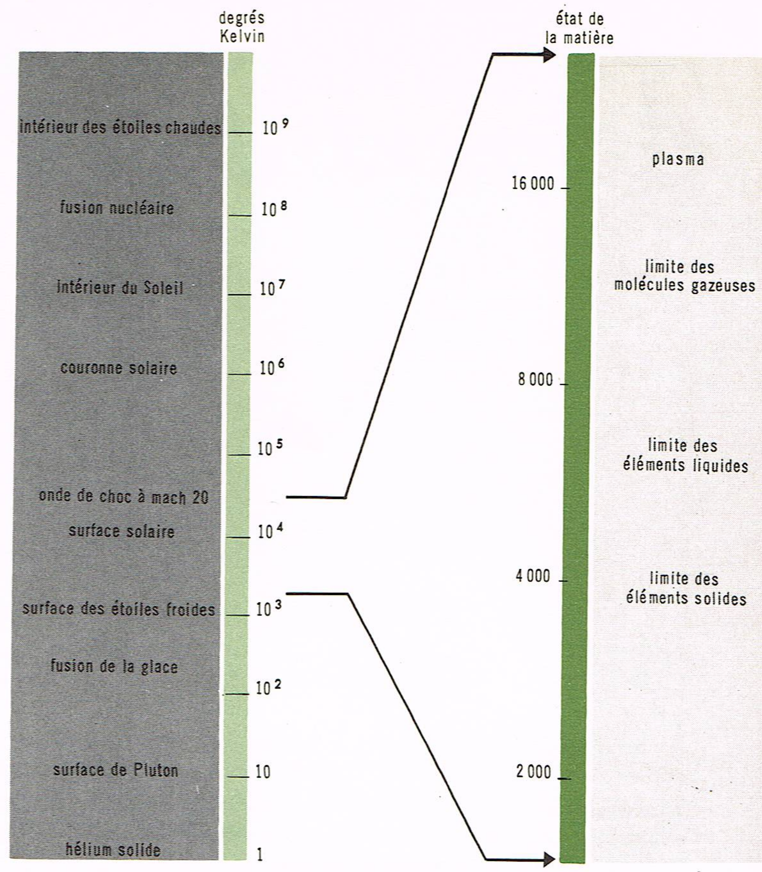
La définition de la température, variable de qualité dans l'expression de la chaleur, se déduit simplement des notions d'énergie et de désordre que nous venons d'introduire. La température thermodynamique d'un système sera la quantité de chaleur qu'il faudra lui fournir pour augmenter son entropie d'une unité. Elle s'exprime en degrés Kelvin (°K). Nous noterons que cette définition donne une notion de la température indépendante des propriétés particulières des appareils qui sont utilisés en pratique pour la mesurer. Le thermomètre à mercure, par exemple, mesure la température en termes d'une propriété spécifique du mercure, sa dilatation, sur une échelle arbitrairement divisée, entre la température de la glace fondante et celle de l'eau bouillante, en 100 unités dites « degrés Celsius (°C)». On démontre que l'échelle des températures thermodynamiques s'identifie à l'échelle des températures absolues telle que la définit un thermomètre à gaz parfait. Cette échelle est limitée inférieurement : elle s'arrête au 0 absolu (0 °K = — 273 °C).

Au contraire, il n'existe pas de limite supérieure à l'échelle des températures, et nous ne rencontrerons donc aucun obstacle théorique à la réalisation de températures de plus en plus élevées.

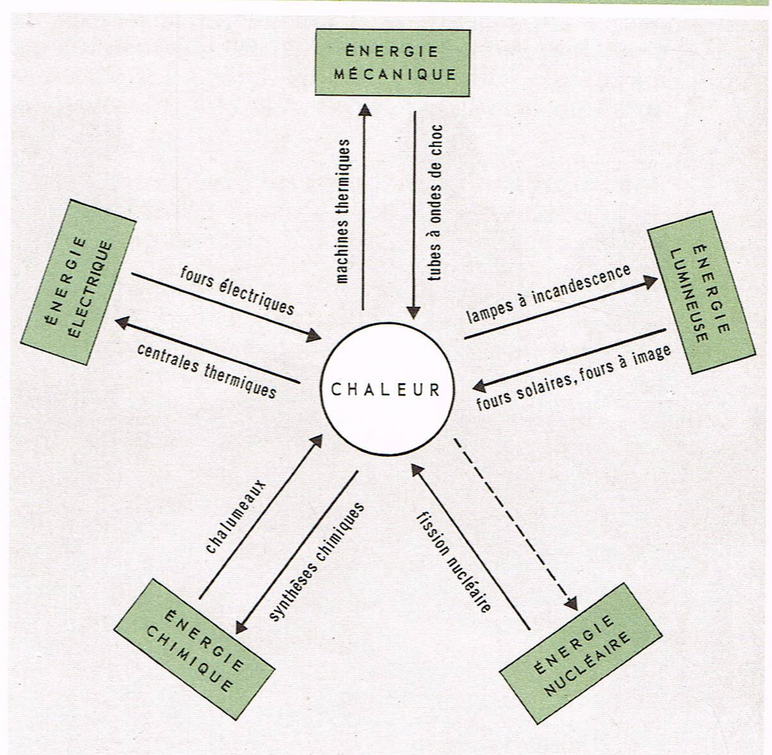
La chaleur est une forme dégradée de l'énergie. Cette troisième proposition constitue en fait le second principe de la thermodynamique et limite notre optimisme sur la rentabilité des transformations de la chaleur : la chaleur est transformable en d'autres formes d'énergie, mais avec un rendement faible. Par exemple, au cours de la transformation, dans une centrale thermique, de l'énergie calorifique en énergie électrique, les deux tiers environ de la chaleur fournie par le combustible sont éliminés par l'eau de refroidissement. Cela explique d'ailleurs le prix élevé de l'énergie électrique d'origine thermique. Au contraire, l'énergie électrique peut être transformée en chaleur sans perte importante.

Du point de vue théorique, et du point de vue théorique seulement, nous serons donc placés pour la production de hautes températures dans des conditions extrêmement favorables, puisque d'une part il n'existe aucune limite théorique à la production de températures de plus en plus élevées, et que, d'autre part, toutes les formes d'énergie pourront être transformées en chaleur avec de hauts rendements.

Le point de vue du réalisateur sera sensiblement différent. L'homme préhistorique a eu besoin de températures élevées



Les températures dans la nature et les états de la matière.



Chaque transformation d'une forme d'énergie en chaleur correspond à un mode de production de hautes températures. Chaque transformation de la chaleur en une autre forme d'énergie correspond à une application des hautes températures.

pour réaliser la cuisson de ses poteries. Inversement, aujourd'hui, si des températures de plus en plus élevées ont fait leur apparition dans les diverses techniques industrielles, il a fallu préalablement mettre au point toute une série de matériaux propres à supporter ces températures. C'est ainsi qu'est née la famille des matières réfractaires. **En fait, dans la plupart des industries mettant en jeu de hautes températures, l'absence de matériaux convenables a été le principal facteur limitant les progrès.** Il ne saurait donc être question de séparer l'étude des matériaux réfractaires de celle des méthodes de production des hautes températures et de leurs applications.

les flammes

(énergie chimique → chaleur)

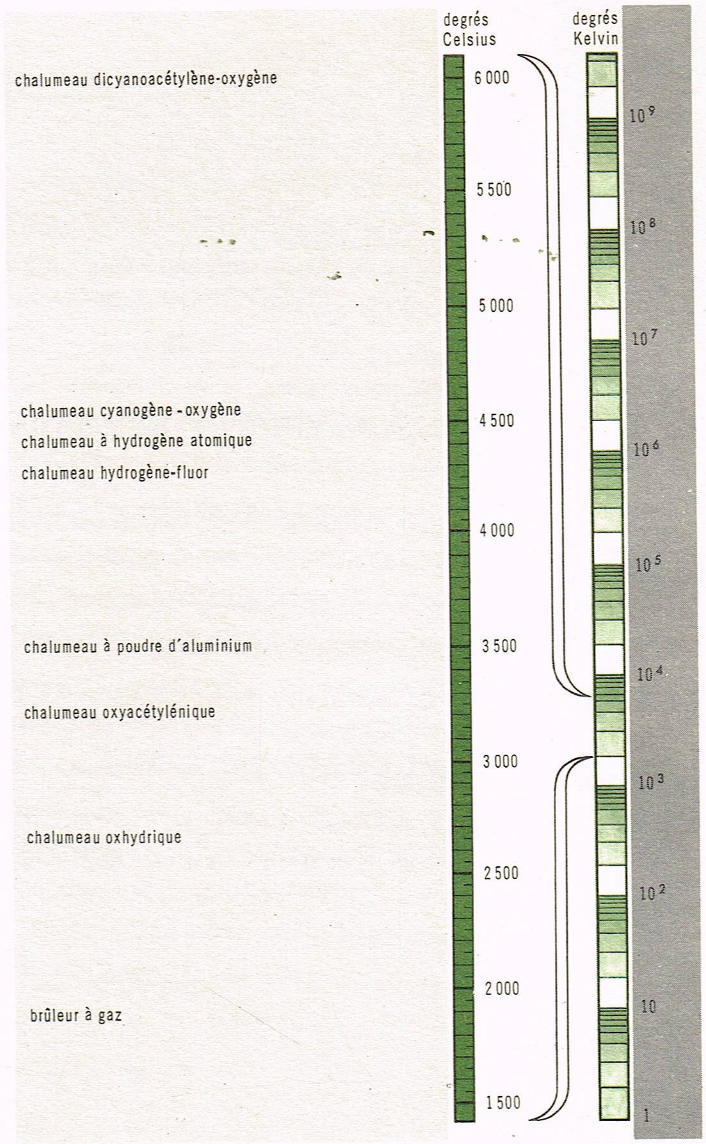
Le feu constitue sans aucun doute le moyen de production de hautes températures le plus anciennement connu. Son histoire est intimement liée à celle de l'humanité. Il est pourtant, même à l'heure actuelle, peu de phénomènes aussi mal connus que le feu et ses manifestations : les flammes.

Si grande fut pour l'homme l'importance du feu qu'il ne se contenta pas de l'utiliser, mais chercha toujours à l'inclure dans une vue philosophique du monde. Aux diverses étapes de la connaissance, jusqu'à Lavoisier, le feu occupera toujours une position particulière dans les systèmes envisagés. Dix-huit siècles après la théorie des quatre éléments d'Aristote, celle du « phlogistique » conserve au phénomène de la flamme un caractère mystérieux d'élément mal défini, dont Lavoisier lui-même ne parvient pas à s'affranchir totalement. C'est cependant Lavoisier qui donne l'explication primaire de la flamme, en montrant le caractère de réaction chimique du phénomène de combustion. C'est le point de départ de la chimie moderne.

Les flammes ne sont autre chose que le résultat d'une combustion, c'est-à-dire d'une réaction chimique entre un combustible et un comburant. Remarquons dès maintenant que notre définition s'est notablement élargie depuis Lavoisier. D'une part, elle ne fait pas nécessairement intervenir l'oxygène comme comburant : certaines substances brûlent dans l'azote, d'autres dans le fluor. D'autre part, elle ne fait pas intervenir la notion de température. En fait, la température des flammes peut varier dans des limites très larges. Le borure d'éthyle s'enflamme spontanément dans l'air, même raréfié, en donnant naissance à une flamme dont la température n'excède guère 100 °C; il s'agit d'une flamme froide. Au contraire, la flamme du chalumeau oxhydrique d'Henri Sainte-Claire Deville atteint 2 500 °C Celle du chalumeau oxyacétylénique, dépassant 3 000 °C, était même jusqu'à ces dernières années l'une des sources les plus chaudes connues.

Mais le problème qui se pose à nous aujourd'hui est celui de l'obtention de températures plus élevées encore. Les méthodes chimiques nous permettront-elles de parvenir à ce résultat, et comment?

Qu'est-ce que la température d'une flamme, et de quoi dépend-elle? On convient d'appeler « température » d'une flamme la température maximale à laquelle sont portés les gaz brûlés. Elle dépend avant tout, naturellement, de la chaleur dégagée au cours de la réaction chimique de combustion et de la chaleur spécifique des gaz formés, c'est-à-dire de la quantité de chaleur qu'il faut fournir à ces gaz pour élever leur température de un degré. Mais, fréquemment, ces gaz formés (gaz carbonique, vapeur d'eau) sont susceptibles de se décomposer à haute température suivant une réaction dite « endothermique », c'est-à-dire absorbant de la chaleur. La chaleur dégagée au cours de la réaction de combustion ne sera donc pas utilisée uniquement pour élever la température de la flamme, mais une partie sera consommée par la réaction de décomposition des gaz brûlés.

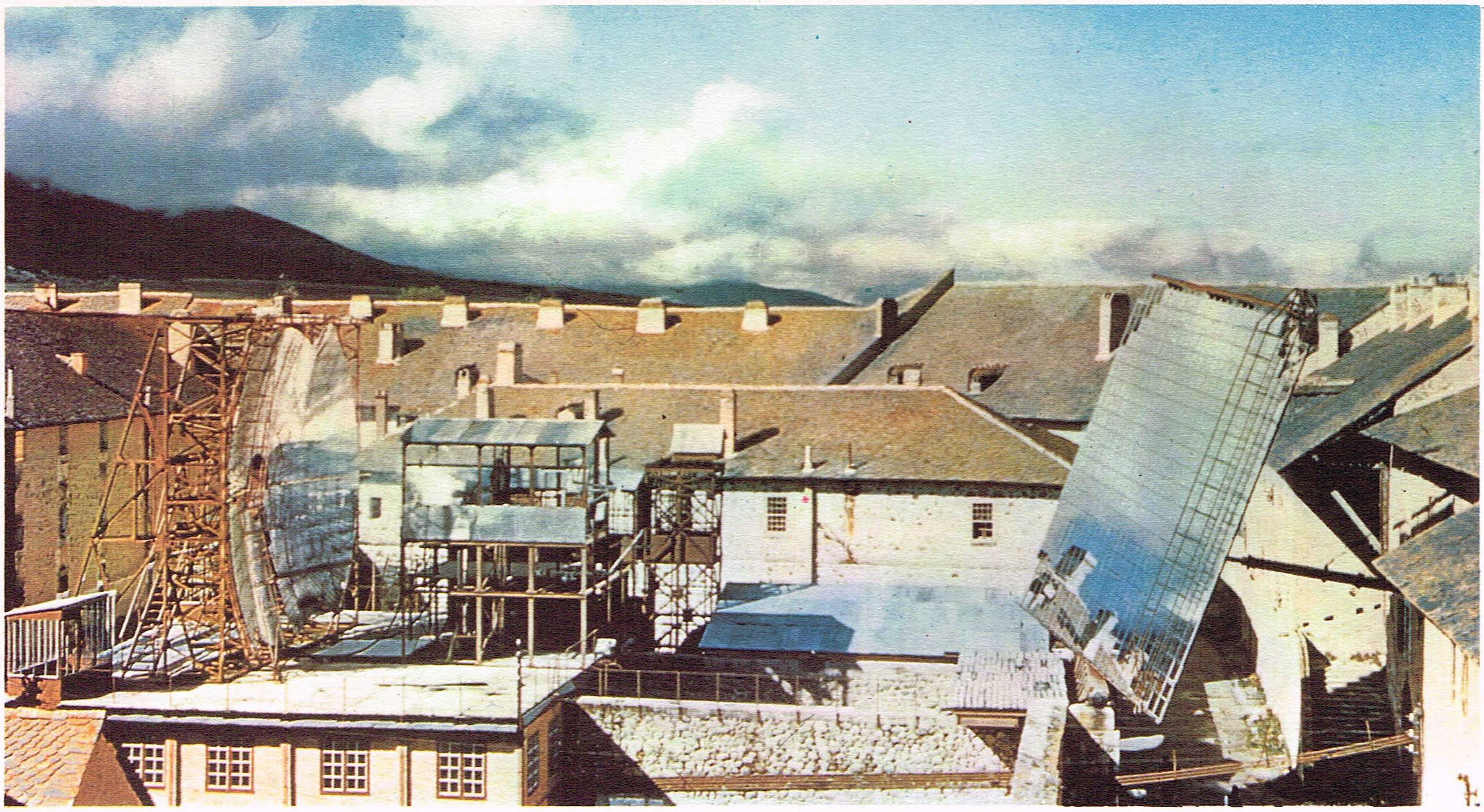


Les flammes les plus chaudes.

Or, bien souvent ce sera finalement ce phénomène de décomposition endothermique qui limitera la température de la flamme.

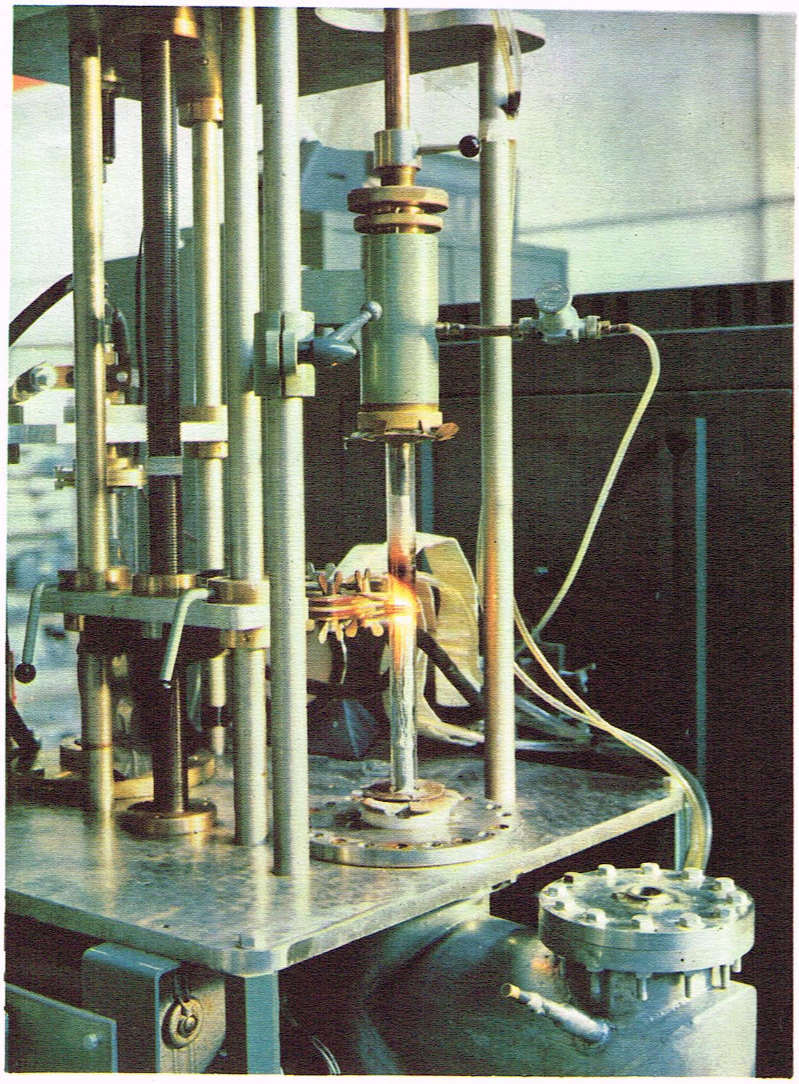
Comment dès lors choisir les éléments de la réaction chimique pour obtenir la plus haute température possible? Il faudra faire appel à des réactions donnant naissance à des corps peu dissociables. Ainsi, nous devons éviter d'associer à l'oxygène un combustible hydrogéné ou carburé : la vapeur d'eau ou le gaz carbonique formés au cours de la réaction sont des corps éminemment dissociables. Le couple combustible-comburant idéal ne sera donc pas le couple traditionnel hydrocarbure-oxygène. La simple comparaison des températures de dissociation montre que les molécules gazeuses les plus stables sont celles de l'acide fluorhydrique, de l'oxyde de carbone, de l'azote et de l'hydrogène. Cette remarque a été à l'origine du développement des chalumeaux modernes, producteurs de très hautes températures.

Né il y a une dizaine d'années aux États-Unis, le chalumeau hydrogène-fluor est fondé sur la réaction de formation de l'acide fluorhydrique : $H_2 + F_2 \rightarrow 2\,HF$. La température obtenue est de l'ordre de 4 100 °C.



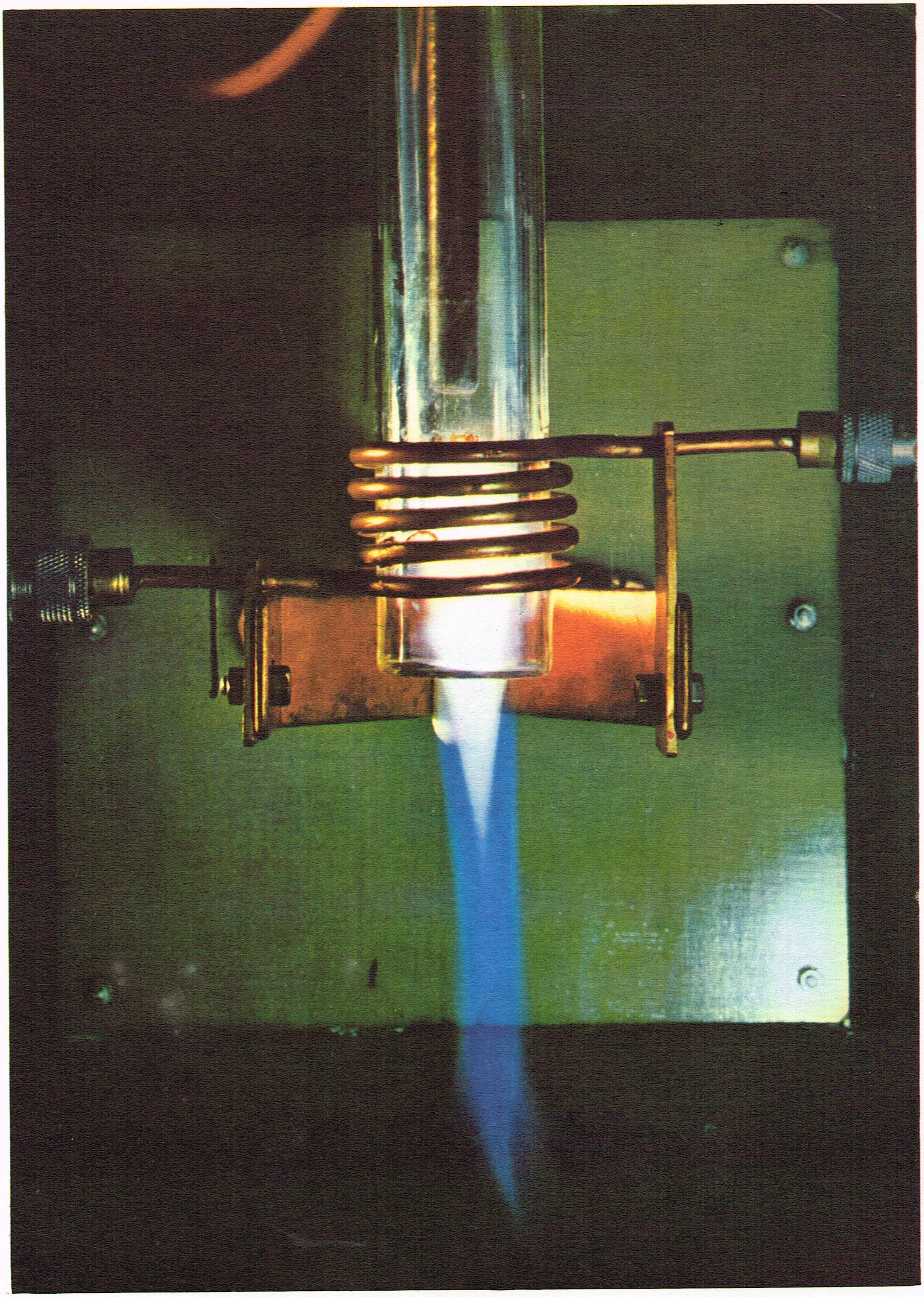
Vue d'ensemble du grand four solaire de Mont-Louis. On distingue l'héliostat de 135 m² *(à droite)*, le miroir parabolique de 90 m² *(à gauche)* et, entre les deux, le laboratoire où sont traitées les substances placées au foyer du miroir.

Doc. C. N. R. S.



Purification du béryllium
dans un four à induction
au Centre d'études nucléaires de Grenoble.

Phot. C. S. E., René Bouillot.



Le beau panache de flammes provoqué par l'injection de zircone dans le four à plasma.
Phot. C. S. F., René Bouillot.

La combustion dans l'oxygène de composés cyanés, c'est-à-dire contenant du carbone et de l'azote, donnera naissance aux molécules très stables d'oxyde de carbone et d'azote. On pourra donc espérer, à l'aide de cette réaction, obtenir des températures extrêmement élevées. L'expérience montre en effet que la flamme du chalumeau cyanogène-oxygène atteint 5 000 °C. Mais la flamme la plus chaude actuellement connue est celle du chalumeau dicyanoacétylène-oxygène ($C_4N_2 + O_2$). Sa température dépasse 6 000 °C, et, au cours des toutes dernières années, la substitution totale ou partielle d'ozone à l'oxygène a même permis de porter la température de cette flamme jusqu'à 7 000 °C.

Autre chalumeau de structure sensiblement différente, mais fondé sur le même principe, le chalumeau à hydrogène atomique, ou chalumeau de Langmuir, permet de réaliser des températures supérieures à 4 200 °C. L'hydrogène moléculaire traverse un arc électrique dans lequel il se dissocie, du moins partiellement, en atomes. Hors de l'arc, ces atomes se recombinent suivant une réaction fortement exothermique : $H + H \rightarrow H_2$.

Avec ses 3 100 °C, le chalumeau oxyacétylénique était considéré il y a seulement une dizaine d'années comme l'instrument des plus hautes températures. Ses performances ne constituent plus aujourd'hui que la première étape dans la conquête d'une nouvelle gamme de températures encore beaucoup plus élevées, celles qui s'étendent de 3 000 à 7 000 °C.

Il serait injuste d'oublier, dans cette analyse des méthodes chimiques de production de très hautes températures, une classe très particulière d'appareils : les chalumeaux à combustion de métaux. Notre choix du métal combustible sera guidé par les mêmes principes qui nous ont conduits dans le choix du gaz.

Le métal devra avoir une forte chaleur de combustion dans l'oxygène, mais également donner naissance à des oxydes stables à haute température. Le premier impératif va nous orienter vers les métaux alcalins (lithium, sodium, potassium), alcalino-terreux (calcium, baryum), le magnésium, l'aluminium, le zirconium.

Le deuxième impératif va éliminer de cette liste les alcalins, à l'exception du lithium, et les alcalino-terreux, dont les oxydes sont peu stables à haute température. Finalement, les deux réalisations les plus intéressantes sont le chalumeau à lithium liquide et le chalumeau à poudre d'aluminium.

Dans le premier de ces deux appareils, les gouttelettes de lithium sont entraînées par un gaz inerte, de l'argon par exemple, et brûlent dans l'oxygène, l'ozone ou le fluor.

Dans le deuxième appareil, les grains de poudre d'aluminium sont entraînés par un courant d'oxygène. La longueur de la flamme est de l'ordre de cinq mètres et sa température de 3 500 °C.

Ce dernier a déjà reçu des applications industrielles. L'une d'elles est le découpage des blocs de béton épais de plusieurs dizaines de centimètres. Une autre application est la réalisation de revêtements d'alumine de quelques millimètres à quelques centimètres d'épaisseur sur la surface des métaux, des matières plastiques ou des réfractaires. C'est ainsi que des briques en argile bon marché acquièrent les qualités réfractaires des briques d'alumine lorsqu'elles sont recouvertes d'une couche d'alumine de quelques centimètres d'épaisseur.

les fours à concentration de rayonnement
(énergie lumineuse → chaleur)

Si, pendant de nombreux siècles, la production de hautes températures a été liée à l'art du feu et a reposé donc sur des méthodes chimiques, d'autres méthodes ont fait depuis longtemps une timide apparition. Tout naturellement, les hommes s'étaient tournés vers leur ami le plus permanent, le plus puissant et le plus efficace : le soleil. Au lieu de se contenter de ses bienfaits naturellement prodigués, ils avaient songé à mieux utiliser ses rayons pour en tirer de hautes températures.

215 ans avant notre ère, Archimède réalisait une première application de transformation d'énergie lumineuse en chaleur : à l'aide d'un grand miroir hexagonal, il incendiait la flotte romaine qui assiégeait Syracuse. Huit siècles après, Procus, à Constantinople, réalisait un exploit du même genre en incendiant la flotte de Vitalien. Mais déjà bien avant, plus pacifiquement, le feu sacré de Vesta était ranimé en tournant vers le soleil la surface concave d'un vase d'or poli. Puis, on semble renoncer à produire des hautes températures par des méthodes de ce genre, et il faut attendre la fin du XVII[e] siècle et les débuts du XVIII[e] pour voir apparaître les miroirs et les verres ardents. Le plus célèbre d'entre eux est le miroir offert par Cassini à Louis XIV, qui permettait de fondre aisément le fer. A la même époque, Buffon, reprenant sur une échelle plus modeste l'expérience d'Archimède, incendiait un tas de bois à 65 m, en utilisant une série de petits miroirs.

Mais c'est à Lavoisier, une fois encore, que revient le mérite d'avoir introduit la rigueur scientifique dans ces essais désordonnés et d'avoir montré tout le parti que l'on pouvait tirer du rayonnement solaire pour le chauffage ou la fusion des corps. Puis c'est de nouveau 170 ans de silence presque total.

Il est vrai que, entre-temps, la découverte de formes d'énergie nouvelles, pétrole et électricité, avait capté l'attention des chercheurs et mis les hommes à l'abri du besoin. C'est en 1946 qu'une équipe de chercheurs français, dirigée par Félix Trombe, reprenait les essais de chauffage solaire à l'observatoire de Meudon.

C'est alors que l'homme, ayant gaspillé une grande partie de ses ressources en énergie, commence à envisager des réalisations industrielles fondées sur l'énergie solaire. C'est à ce besoin que répondent le prototype industriel d'une puissance de 80 kW, créé en 1952 par le C. N. R. S., à Mont-Louis, et surtout le très grand four de 1 000 kW en cours de construction à Odeillo.

Ces magnifiques réalisations ont été décrites dans plusieurs ouvrages de Félix Trombe (v. notamment *les Grandes Découvertes du XX[e] siècle*, Larousse), dont on ne saurait trop recommander la lecture.

Mais l'utilisateur du four solaire est évidemment tributaire de l'éclat du soleil, et l'on sait bien à quel point certaines années peuvent être déficitaires en ensoleillement. Il est aussi tributaire de la transparence de l'atmosphère, et Pol Duwez, réalisateur en 1937 de l'un des tout premiers fours solaires à Pasadena (Californie), a vu décroître progressivement les possibilités d'utilisation de son appareillage à mesure qu'augmentait ...le nombre de voitures de Los Angeles.

Aussi a-t-on recherché la mise au point d'un mode de chauffage dérivé du chauffage solaire, c'est-à-dire conservant ses immenses avantages (très hautes températures et chauffage des matériaux sans pollution d'aucune sorte), mais utilisable en tout lieu, à toute époque de l'année.

Un tel dispositif existe dans différents pays : c'est le four à image. Son principe est connu depuis fort longtemps, et une gravure ancienne représente le professeur Tyndall démontrant pour la

première fois dans ses leçons sur le rayonnement les possibilités du four à image d'arc.

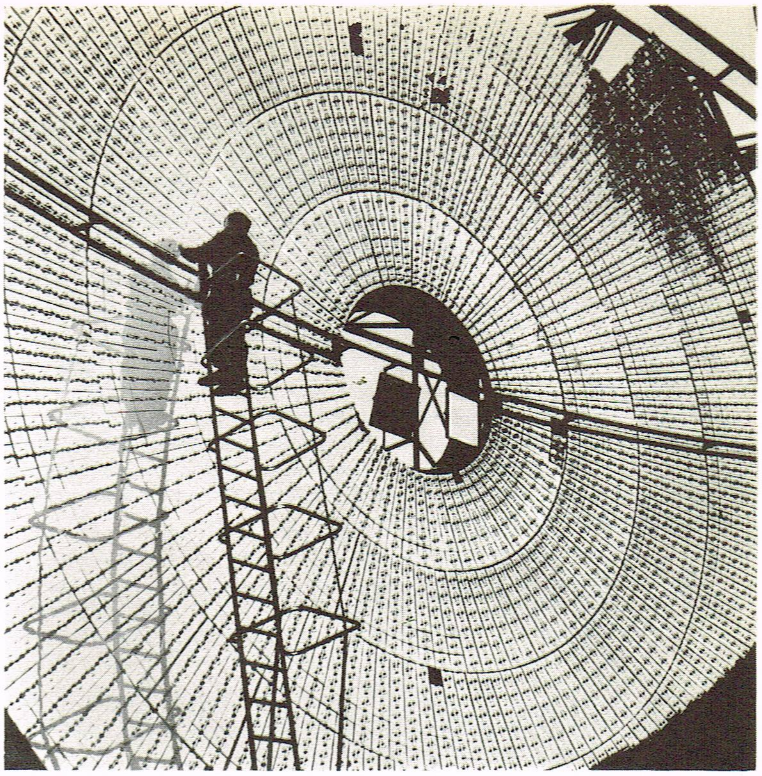
Considérons un corps porté à très haute température, le cratère d'un arc électrique par exemple, et plaçons-le au foyer d'un miroir parabolique M_1. Les rayons lumineux émis par la substance subiront une réflexion sur le miroir et donneront naissance, suivant les propriétés de la parabole, à un faisceau de rayons parallèles. Interposons sur leur trajet un deuxième miroir parabolique M_2, faisant face au premier. Après réflexion, on obtiendra un faisceau venant converger au foyer de ce deuxième miroir. Nous aurons donc finalement transporté la source chaude du foyer A au foyer E. Nous aurons obtenu une image de cette source. Remarquons bien qu'au foyer A nous n'aurions pu effectuer le chauffage d'une substance qu'au contact avec la source chaude, ce qui aurait entraîné une inévitable pollution. En E, au contraire, nous récupérons de la chaleur pure, sans support matériel. **Il en résulte que, quelle que soit la nature de la source chaude, nous pourrons chauffer en E n'importe quelle substance, si réactive soit-elle.**

L'utilisateur du four à image dispose donc d'une totale liberté, et son imagination peut se donner libre cours pour choisir aussi bien la nature de la source chaude que celle du système optique. La source peut être une flamme, par exemple celle d'un chalumeau à combustion d'aluminium. Mais il est bien évident que l'énergie émise est alors répartie sur toute la longueur de la flamme et qu'un faible pourcentage se retrouvera sur l'image ponctuelle. Aussi, la meilleure solution, le plus souvent adoptée, consistera-t-elle à choisir comme source le cratère du charbon d'un arc électrique.

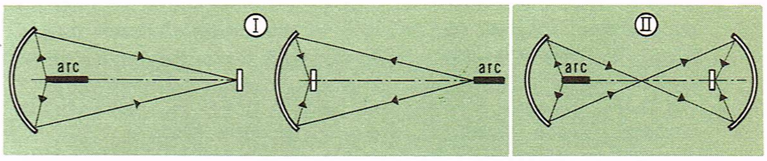
Le système optique peut, lui aussi, être très varié. Il sera constitué, comme dans l'exemple choisi, par deux miroirs paraboliques placés face à face (type III). On pourra aussi mettre à profit les propriétés de l'ellipse. Supposons une source chaude placée à l'un des foyers d'un ellipsoïde de révolution. Après réflexion sur la surface polie de l'ellipsoïde le rayonnement viendra converger au second foyer. Dans ce dispositif, le rayonnement subit donc une réflexion unique. D'autre part, on peut obtenir, suivant la position de l'échantillon à l'un ou l'autre foyer, une zone étendue à faible éclairement énergétique ou une zone réduite subissant une forte irradiation (type I).

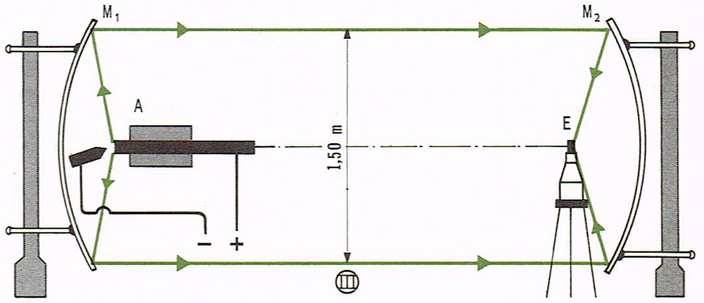
On peut encore augmenter la souplesse de l'appareillage en utilisant le four dit « à ellipsoïdes conjugués » (type II). La source chaude (arc) est placée au foyer F_1 d'un premier miroir elliptique. Après réflexion, le rayonnement vient converger au second foyer F'_1. Cette image va servir de source à un second miroir elliptique, dont le foyer F_2 sera confondu avec F'_1. La substance à chauffer sera placée en F'_2. La présence du foyer intermédiaire F'_1 F_2 permettra aisément d'interposer des écrans ou des miroirs et de modifier ainsi l'orientation du rayonnement.

Quelles températures peut-on atteindre à l'aide de ces fours à image? La température de l'image sera naturellement au maximum égale à celle de la source. Avec un arc électrique ordinaire, elle ne dépassera guère 3 200 °C, mais avec certains types d'arcs dits « arcs intensifs », elle pourra atteindre 4 700 °C. Remarquons cependant que la température obtenue à l'aide d'un tel système dépendra bien moins de l'énergie fournie par la source lumineuse que des caractéristiques du corps à chauffer : pouvoir absorbant, pouvoir émissif, pouvoir réflecteur, etc. Cependant, tous les matériaux, même les plus réfractaires, ont pu être fondus au foyer du four à image. Bien d'autres techniques conduiraient au même résultat. Mais il convient surtout de remarquer une nouvelle fois que le four à image constitue l'une des rares méthodes permettant de réaliser le chauffage de matériaux quelconques à de très hautes températures dans des conditions de propreté rigoureusement parfaites, puisque la chaleur est transmise au matériau à chauffer sans aucun intermédiaire matériel. Retenons, d'autre part, que

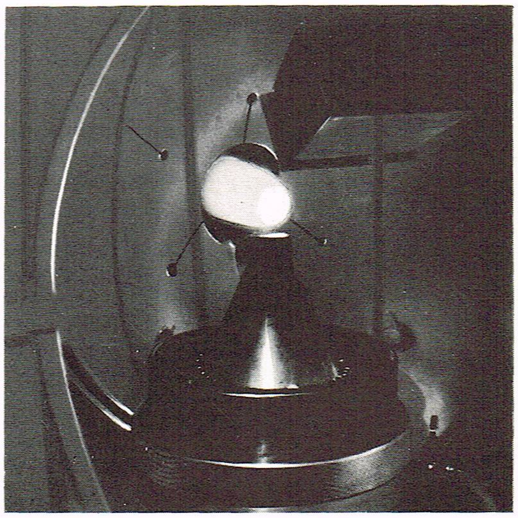


Miroir parabolique de 90 m² du four solaire de 80 kW installé à Mont-Louis. On distingue les couronnes concentriques constituées d'éléments identiques.





Principe de fonctionnement d'un four à image à double miroir parabolique.



Four à image : la source chaude (le cratère du charbon de l'arc électrique) est « transportée » à l'aide du système optique sur le support refroidi à l'emplacement de l'objet à chauffer.

Phot. C. S. F., René Bouillot.

le chauffage peut être réalisé à l'intérieur d'enceintes transparentes et dans n'importe quelle atmosphère.

L'absence totale de pollution et l'extrême souplesse de fonctionnement constituent des avantages fondamentaux du four à concentration de rayonnement.

Aussi voit-on aujourd'hui se développer les applications des fours à image dans les domaines les plus divers.

C'est sans aucun doute dans les domaines de la chimie et de la métallurgie que le four à image a rendu les plus grands services. Mais il faut noter dès maintenant qu'il a fait aussi son apparition dans les laboratoires de physique et, assez curieusement, dans les laboratoires de biologie. Le four à image s'est révélé en effet comme l'instrument susceptible de reproduire le plus fidèlement possible les brûlures cutanées provoquées par les explosions nucléaires.

les fours électriques
(énergie électrique → chaleur)

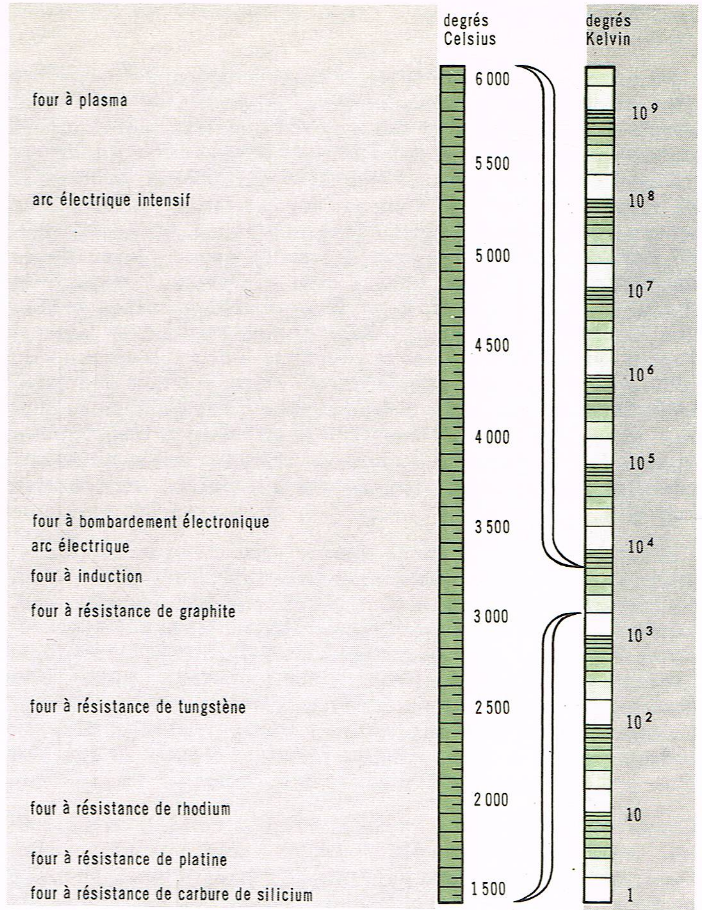
Dans la conquête du domaine des hautes températures, le pas décisif fut franchi lorsque le physicien anglais Davy découvrit l'arc électrique au début du XIX[e] siècle.

L'étonnante souplesse de l'énergie électrique a donné naissance à une foule d'appareils de principes très différents. Il ne saurait être question de les énumérer ici, et nous nous bornerons à citer quelques-unes des températures records réalisées.

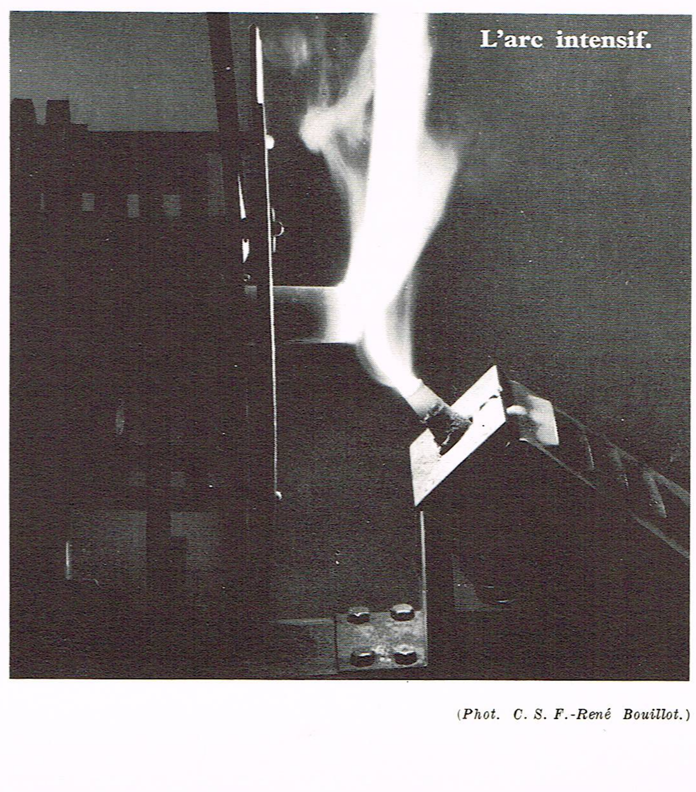
L'effet Joule fut l'une des premières manifestations bien connues du courant électrique. Aussi, les fours à résistances furent-ils les premiers utilisés. Ces fours diffèrent essentiellement par la nature de l'élément conducteur, qui varie suivant la température à atteindre, et l'atmosphère du four : à l'air ou en atmosphère oxydante, alliages inoxydables jusqu'à 1 300 °C, carbure de silicium jusqu'à 1 500 °C, métaux précieux jusqu'à 1 600 °C (platine) ou 1 800 °C (rhodium); en atmosphère réductrice ou neutre, molybdène jusqu'à 2 100 °C, tungstène jusqu'à 2 600 °C ou graphite jusqu'à 2 800 °C.

Mais, c'est grâce au four à arc que les températures vraiment élevées ont pu être réalisées. C'est à Henri Moissan que revient le mérite d'avoir montré, à la fin du siècle dernier, tout le parti que l'on pouvait tirer de la puissance calorifique de l'arc pour l'élaboration de nouvelles substances minérales. Les découvertes successives de Moissan engendrèrent, dès les premiers mémoires, un rapide essor des industries électrothermiques. Aujourd'hui, les fours à arc industriels atteignent des dimensions et des puissances colossales, de 30 000 à 50 000 fois supérieures à celles qu'utilisait Moissan. Les performances du four à arc ont été encore améliorées récemment, après l'apparition de l'arc intensif. Le principe de cet appareil consiste à augmenter dans des proportions considérables la densité du courant, par la création d'une zone fortement conductrice au voisinage de l'anode. Ce résultat est obtenu par la volatilisation de minéralisants divers, tels que les oxydes ou les fluorures des terres rares, incorporés à l'anode lors de sa fabrication. La température atteinte dans le cratère de l'arc peut alors dépasser 5 000 °C.

Dernier-né des fours électriques industriels, le four à induction réalise le chauffage des matériaux métalliques dans les meilleures conditions, puisque la chaleur est produite au sein de la substance elle-même. Tout risque de pollution par les parois du four est ainsi éliminé. On peut très schématiquement considérer le four comme le secondaire d'un transformateur dont le primaire est parcouru par le courant d'induction. L'industrie métallurgique utilise des fours à induction dont la puissance dépasse aujourd'hui 1 000 kW. Ces fours se prêtent d'une manière parfaite au chauffage soit sous vide, soit sous différentes atmosphères.



Echelle des températures atteintes dans les fours électriques.



L'arc intensif.

(*Phot. C. S. F.-René Bouillot.*)

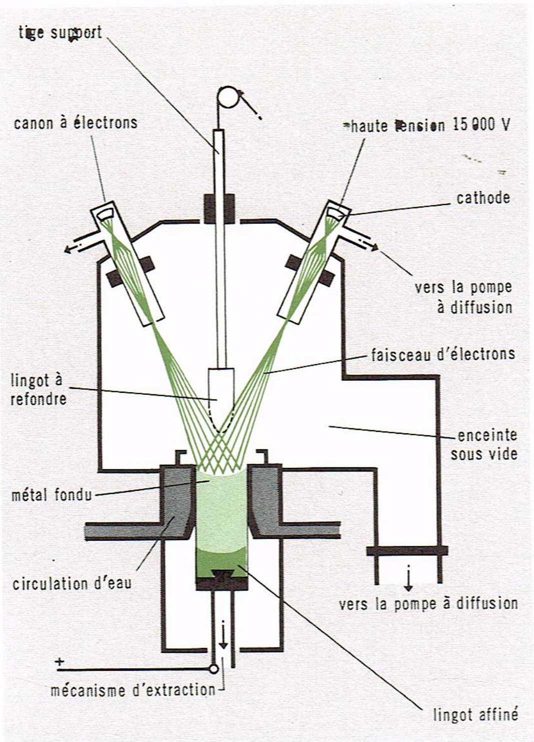
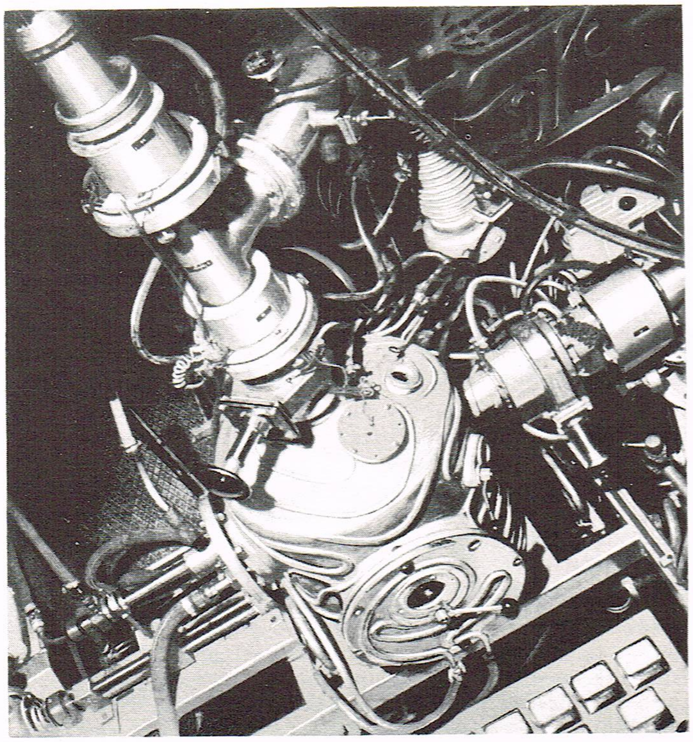


Schéma d'un four de fusion par bombardement d'électrons (canon à faisceau électronique focalisé à « longue distance »).



L'impressionnant ensemble d'un four à bombardement électronique de 20 kW.

Le chauffage par induction présente aussi un caractère tout à fait particulier : la température développée à l'intérieur du matériau dépend bien moins de la puissance du four que des caractéristiques du matériau lui-même, notamment de sa résistivité. D'une manière générale, la fréquence optimale des courants induits sera sensiblement proportionnelle à la résistivité du matériau à chauffer. Grâce à la réalisation de fréquences de plus en plus élevées, les applications du four à induction, limitées il y a quelques années au chauffage métallurgique, s'étendent aujourd'hui au chauffage des semi-conducteurs et des diélectriques.

Mais, à peine le four à haute fréquence venait-il d'apparaître comme l'instrument susceptible de libérer le métallurgiste des servitudes des fours classiques, qu'un nouveau concurrent se manifestait. En fait, ce n'était pas un nouveau venu.

Le four à bombardement électronique possède en effet une curieuse histoire, qui n'est pas sans analogie avec celle du four solaire ou du four à image. Il y a près d'un siècle que Crookes fondait pour la première fois du platine dans un faisceau d'électrons. Mais c'est seulement en 1934 que Félix Trombe entreprit réellement d'appliquer pour la première fois le four à électrons à l'étude d'un problème pratique : l'isolement et la purification des métaux des terres rares. Puis l'intérêt porté à cet appareillage faiblit notablement jusqu'à ces dernières années, où l'on redécouvrit pour la troisième fois le four à électrons.

Le principe du four à électrons est très simple; c'est celui même des tubes à rayons X, mais sa réalisation est fort délicate. Un filament métallique chauffé sous vide émet des électrons. Ceux-ci, accélérés par une différence de potentiel convenable, viennent heurter un obstacle et lui communiquent sous forme de chaleur leur énergie cinétique.

Les températures atteintes dépassent 3 500 °C et permettent la fusion de tous les métaux, y compris le tungstène (près de 3 400 °C). Des fours de 250 kW sont dès maintenant en service, et des fours de 1 500 kW sont actuellement à l'étude.

C'est le métallurgiste qui observe avec le plus d'intérêt le développement de ces fours. Il en attend essentiellement l'élimination de la pollution des métaux par les creusets, la possibilité d'obtenir des températures de plus en plus élevées par la concentration localisée de très grandes puissances, celle de faire varier dans de larges limites la vitesse de fusion, de traiter sous vide très élevé les métaux fondus pour produire l'évaporation des impuretés.

Nous aurions terminé cette analyse fort incomplète des méthodes de transformation d'énergie électrique en chaleur, si un nouvel appareil n'avait fait ces dernières années une fracassante apparition, le four à plasma, qui constitue le plus puissant moyen de chauffage contrôlable actuellement connu. Son importance est telle, aujourd'hui, que nous l'étudierons plus en détail.

les tubes à ondes de choc
(énergie mécanique → chaleur)

La résistance exceptionnelle de l'air au passage d'un corps se déplaçant à une vitesse supérieure à celle du son est remarquée pour la première fois en 1742, au cours d'essais d'artillerie. Près de 150 ans devaient s'écouler avant que l'on reconnût que l'accroissement de cette résistance était dû à la formation d'ondes intenses de compression de l'air au voisinage du corps : les ondes de choc. Or, par la compression d'un gaz, on réalise une augmentation de sa température. Si la compression est appliquée à grande vitesse, par exemple si elle est engendrée par une explosion ou par un avion supersonique, la plus grande partie de l'énergie mécanique est transformée en chaleur. Si le nez de l'avion supersonique n'était refroidi, sa température

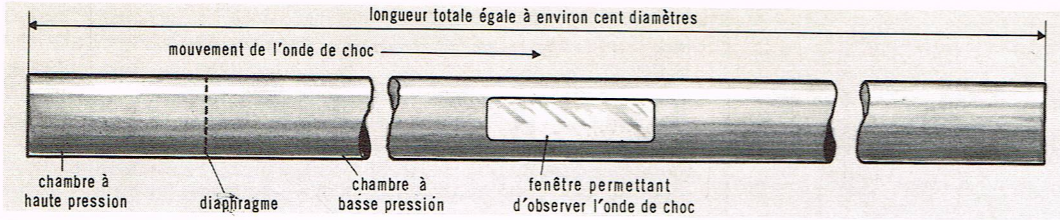


Schéma d'un tube à ondes de choc.

s'élèverait à 1 000 °C pour mach 4, à 3 000 °C pour mach 10 et à 20 000 °C pour mach 20.

On a pu donc se demander si l'utilisation d'une onde de choc puissante, mais contrôlée, ne pourrait pas constituer une méthode de production de très hautes températures. De cette idée est né le tube à ondes de choc. C'est un long tube d'acier de 30 m de long et 30 cm de diamètre environ, séparé en deux parties par un épais diaphragme de cuivre. La partie 1 est remplie d'hydrogène, la partie 2 d'argon sous pression réduite. L'explosion de l'hydrogène auquel on ajoute un peu d'oxygène est provoquée par une étincelle. Le diaphragme se brise alors et l'onde de choc se propage dans l'argon. Les vitesses obtenues peuvent atteindre mach 20, et des températures de l'ordre de 20 000 °C ont ainsi pu être réalisées.

Ces techniques très spécialisées semblaient être, jusqu'à maintenant, réservées à l'expérimentation en laboratoire et non à la production industrielle. Cependant, la production d'ondes de choc par l'explosion d'une série de charges chimiques judicieusement disposées pourrait constituer un procédé relativement économique d'obtention de très hautes pressions et de très hautes températures simultanément. Ne vient-on pas d'apprendre récemment que la synthèse du diamant avait été réalisée en plaçant convenablement des cristaux de graphite dans l'onde de choc produite par une explosion chimique. Les rendements de conversion du graphite en diamant ne dépassent pas encore 5 p. 100 et les cristaux produits ont des dimensions inférieures à 40 μ, mais il n'est pas impossible que ce procédé très récent soit finalement plus économique que la technique actuelle, fondée sur l'utilisation de fortes pressions statiques.

les explosions nucléaires *(énergie nucléaire → chaleur)*

Nous atteignons ici le domaine des plus hautes températures. Ce sont celles qui règnent à l'intérieur des plus chaudes étoiles. Elles sont caractérisées par leur aptitude à entretenir les réactions thermonucléaires. Ces réactions appartiennent à deux groupes principaux : aux environs de 15 millions de degrés, elles intéressent surtout les noyaux des atomes légers (cycle de Bethe); au-dessus de 100 millions de degrés, elles intéressent également les atomes plus lourds (cycle du carbone).

Il est bien évident que le problème de l'utilisation de telles températures ne se pose même pas encore.

Des projets américains très sérieux ont cependant proposé récemment d'utiliser des explosions thermonucléaires souterraines pour l'exploitation de mines de sel au Nouveau-Mexique. Une unique détonation de l'ordre de la mégatonne fournirait 3 millions de tonnes de sel fondu, soit une masse sphérique de 200 m de diamètre.

On pourrait également considérer une explosion nucléaire souterraine comme un énorme réacteur chimique à très haute température. En fait, une première explosion souterraine, dont les résultats sont aujourd'hui bien connus, a eu lieu en septembre 1957. Elle a mis en œuvre une charge nucléaire correspondant à 1 700 tonnes de T. N. T. sous une colline dans l'Etat de Nevada. On a pu constater que la chambre d'explosion avait été transformée en une vaste cavité sphérique de 40 m de diamètre. Les dimensions de cette grotte artificielle montraient que 700 t de roches au moins avaient été complètement vaporisées. Les parois étaient faites de roches vitrifiées formant une coquille brillante de 10 cm d'épaisseur. Cette lave avait coulé en se solidifiant, formant des stalactites et des stalagmites. Des fragments examinés au laboratoire montrèrent, comme on pouvait s'y attendre, que des pressions fantastiques s'étaient exercées dans la cavité. Les fragments une fois chauffés se dilataient de plusieurs fois leur volume, ce qui indiquait que la pression de pointe devait avoisiner 6 millions d'atmosphères.

l'accès au domaine des températures supérieures à 5000 °C

Le rapide survol des méthodes de production de hautes températures auquel nous venons de nous livrer nous aura permis d'apercevoir dans l'échelle des températures produites une énorme lacune.

D'une part, les fours classiques assurent assez bien le chauffage dans un domaine s'étendant de la température ordinaire à 6 000 °C environ. D'autre part, le développement des technologies nucléaires durant les dix dernières années nous a habitués à la notion des températures de l'ordre de plusieurs millions de degrés.

Entre ces deux limites s'étend un immense domaine à peu près encore inexploré tant du point de vue des réactions chimiques que du point de vue des propriétés des corps. Au cours des dernières années, quelques incursions ont été tentées grâce aux tubes à ondes de choc, par exemple, mais aucune réalisation pratique et surtout industrielle ne pouvait être envisagée.

Or, voici que, après quelques tâtonnements de courte durée, un nouveau groupe d'appareils vient d'être mis au point, qui va permettre aux chercheurs d'explorer ce domaine non plus par quelques essais sporadiques, mais d'une manière pratique, continue et très bientôt industrielle.

quelques grandes dates de l'histoire des hautes températures

— 800000 (?) - le feu
— 215 - l'ancêtre des fours solaires. Archimède l'utilise pour incendier la flotte romaine.

les premiers pas

1700-1780 - les miroirs et les verres ardents. Lavoisier les utilise pour fondre le platine.
Lavoisier interprète le phénomène de la combustion

1808 - Davy découvre l'arc électrique
Les principes des grandes méthodes de production de hautes températures sont, dès lors, établis.

les hautes températures

1870 - le chalumeau oxhydrique (Henri Sainte-Claire Deville) : 2 500 °C

1880 - le premier four cathodique (Crookes)

1890 - les fours à résistance de métaux précieux : 1 500 °C

1892 - le four à arc d'Henri Moissan : 3 500 °C
Moissan prépare une foule de composés nouveaux et réussit vraisemblablement la synthèse du diamant.

1905 - les fours à résistance de graphite : 3 000 °C

1915 - les fours à résistance de métaux réfractaires : 2 300 °C

1920 - les fours à haute fréquence

les très hautes températures

1922 - les principes des appareils à plasma (Gerdien)

1925 - le chalumeau à hydrogène atomique (Langmuir) : 4 200 °C

1945 - 20 000 °C dans un tube à ondes de choc

1950 - le chalumeau hydrogène-fluor : 4 100 °C

1951 - le premier chalumeau à plasma-arc (Maeker) : 20 000 °C

les grandes réalisations modernes

1952 - les grands fours solaires (Mont-Louis)

1953 - les fours à image

1955 - les grands fours à bombardement électronique

1955 - Bridgman réussit la synthèse industrielle du diamant

1957 - les chalumeaux et fours à plasma

1960 - le laser

la fusion nucléaire contrôlée

1963 - vers les 100 millions de degrés - la température de 40 millions de degrés est maintenue pendant 10^{-2} seconde

à la manière de Monsieur Jourdain

La curiosité qui a entouré l'apparition de ce nouveau venu est due pour une bonne part à l'étrange nom qui lui a été attribué lors de son baptême : chalumeau à plasma.

Le terme de « plasma » a été introduit par Langmuir pour désigner un milieu ionisé. Langmuir mettait ainsi l'accent sur l'instabilité de ce milieu, les difficultés de son confinement, son caractère « gélatineux ».

Le terme de « plasma » sert donc à désigner un état de la matière dans lequel les molécules ne sont plus le constituant ultime : elles sont dissociées en atomes neutres, en ions et en électrons. En fait, les plasmas sont connus depuis fort longtemps. Ils existent dans toutes les décharges dans les gaz, aussi bien dans un arc électrique que dans un tube fluorescent. De sorte que nous savons tous faire des plasmas à la manière de Monsieur Jourdain. D'autre part, on sait aujourd'hui que les flammes sont essentiellement des milieux ionisés. On peut donc trancher d'une manière définitive le problème d'antériorité dans la découverte des plasmas : l'inventeur du plasma s'appelait Prométhée. Et l'on peut aujourd'hui, à condition de ne pas attacher trop d'importance à cette analogie, faire correspondre aux quatre éléments des Anciens quatre états de la matière : solide (la terre), liquide (l'eau), gazeux (l'air), plasma (le feu).

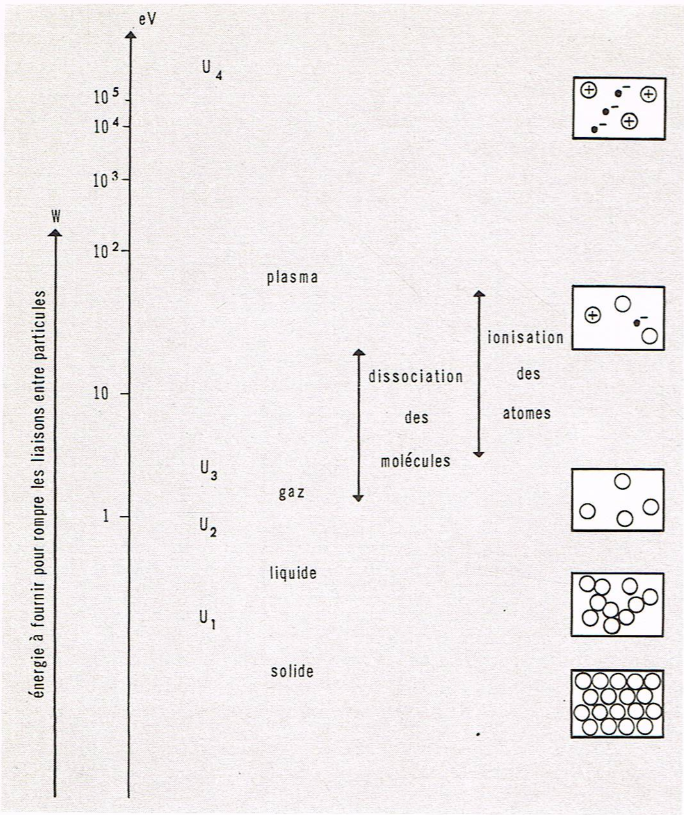
un quatrième état de la matière

Chaque état de la matière correspond, en fait, à un certain degré d'organisation des particules. Cette organisation est déterminée par une valeur assez bien définie de l'énergie de liaison entre ces particules.

Considérons un solide cristallisé. Les molécules qui le constituent sont maintenues entre elles par des forces de liaison, et l'on définit une énergie U_1 des liaisons dans le cristal qui assure à la substance une forme et un volume propres. Chauffons progressivement ce solide. En nous livrant à cette opération, nous communiquons aux particules constituant le cristal une certaine énergie cinétique W. Tant que cette énergie cinétique demeure inférieure à l'énergie de liaison U_1, le corps reste solide. Mais, dès que W devient supérieur à U_1, l'édifice solide va s'écrouler. Un nouvel état va apparaître : l'état liquide, dans lequel la liaison entre particules s'effectuera par des forces de Van der Waals, assurant au liquide un volume propre, mais non une forme déterminée. A ces forces correspond une certaine énergie U_2. Continuons à chauffer notre substance, donc à accroître l'énergie cinétique W. Lorsque W devient supérieur à U_2, un nouvel état plus souple va encore se former : l'état gazeux, où la matière n'aura plus ni forme ni volume propres. Les valeurs de U_1 et de U_2 s'expriment en électron-volts (eV) : elles sont très grossièrement de l'ordre de 1 eV.

L'état gazeux est stable jusqu'à ce que l'énergie cinétique W atteigne une certaine valeur U_3, à partir de laquelle l'élévation de température communique aux particules élémentaires une énergie telle que les liaisons intramoléculaires ou intra-atomiques elles-mêmes sont rompues : des électrons sont alors arrachés à l'atome, en donnant naissance à des ions positifs. L'énergie U_3 est de l'ordre de 10 eV; U_3 est le potentiel d'ionisation de la substance considérée. Le domaine du quatrième état de la matière, le plasma, commence alors. Ce domaine va s'étendre jusqu'au moment où l'atome aura perdu tous ses électrons et où le milieu sera constitué de noyaux privés de leurs cortèges électroniques et d'électrons libres.

Mais W, continuant à croître, va atteindre une valeur U_4 de l'ordre de 10^6 eV, où les liaisons intranucléaires elles-mêmes vont se rompre : les nucléons, particules constituant le noyau, vont être libérés. Ce cinquième état de la matière sera donc un gaz de particules élémentaires. (Rien n'empêche d'ailleurs de prévoir l'existence d'un sixième état, dans lequel ces particules élémentaires en auraient engendré d'autres...)



États de la matière et énergie cinétique.

On tendra donc théoriquement, en élevant la température d'un système, vers des édifices de moins en moins élaborés : molécules, atomes, ions + électrons, noyaux + électrons, nucléons.

Cette représentation est certes fort séduisante, mais elle ne doit cependant pas nous abuser, car, si l'on peut définir avec précision une température de fusion et une température d'ébullition d'une substance sous une pression donnée, il n'existe ni température de dissociation de la molécule, ni température d'ionisation.

D'une part, les domaines de stabilité des états dissociés et ionisés se superposent : les énergies de liaison entre atomes dans la molécule sont en effet du même ordre de grandeur que les énergies d'ionisation. Par exemple, le domaine des énergies d'association entre atomes s'étend de 1 eV pour les molécules organiques les plus faibles à 8 eV pour les molécules les plus solides (azote, oxyde de carbone). Le domaine des énergies d'ionisation s'étend de 3,87 eV pour le césium à 24 eV pour l'hélium. Il n'est pas possible d'établir une correspondance précise entre l'échelle des énergies et l'échelle des températures. On peut dire cependant grossièrement qu'il faut dépasser une température de 8 000 °K pour obtenir une ionisation importante.

D'autre part, dissociation et ionisation existent à toute température et augmentent lorsque la température augmente. L'existence de ce pseudo-quatrième état de la matière correspond donc à une température quelconque, contrairement aux états précédents, à un équilibre entre les particules neutres M (molécules ou atomes) et les particules chargées (ions + électrons) :

$$M \rightleftarrows M^+ + e^-.$$

Dans la majorité des plasmas l'équilibre thermodynamique n'est pas réellement établi. Les tubes fluorescents qui nous éclairent contiennent un plasma dans lequel les électrons sont à des températures de l'ordre de 100 000 °K. Mais, par suite de leur masse très faible, ils sont incapables de transmettre leur énergie aux particules plus grosses (ions et atomes neutres). De sorte qu'il existe en fait dans le tube trois températures distinctes : celle des électrons (de l'ordre de 100 000 °K), celle des ions (de l'ordre de 400 °K), celle des atomes neutres (de l'ordre de 300 °K). Un tel milieu n'est pas en équilibre. Remarquons d'ailleurs que, par suite précisément des différences de masses, la température de l'ensemble ne dépasse guère celle de la paroi. D'une manière générale, l'équilibre sera d'autant plus mal réalisé que l'équipartition de l'énergie se fera plus mal, donc que les collisions entre particules différentes seront plus rares, c'est-à-dire que la pression sera plus basse. Dans un plasma produit sous la pression atmosphérique ou, mieux, sous quelques dizaines ou quelques centaines d'atmosphères, l'équilibre sera sensiblement réalisé.

Le degré d'ionisation et donc la conductibilité du plasma augmentent alors très rapidement avec la température. Toute influence d'origine extérieure, par exemple la présence d'une paroi, aura pour effet de modifier le contour du plasma en modifiant sa conductibilité : une paroi chaude aura tendance à gonfler le plasma, une paroi froide aura tendance à le contracter, donc à accroître le gradient thermique radial et l'effet thermique axial.

C'est sur ce principe que sont construits les appareils générateurs de hautes températures fondés sur les propriétés des plasmas : les chalumeaux à plasma et les fours à plasma.

les chalumeaux à plasma

Dans ces appareils, le plasma est engendré par un arc électrique. Le chalumeau à plasma type est constitué par une chambre généralement refroidie, portant à ses extrémités deux électrodes : une anode généralement creuse, et constituée par une buse de cuivre refroidie, et une cathode faite d'un élément réfractaire (graphite ou tungstène), située à quelques millimètres de l'anode.

L'arc électrique, d'une puissance de quelques kilowatts, est amorcé par un processus quelconque, par exemple par une bobine à haute fréquence. Dès l'amorçage de l'arc, des électrons quittent la cathode et se déplacent vers l'anode d'un mouvement uniformément accéléré. Sur leur trajet, ils rencontrent les atomes du fluide contenu dans la chambre et leur communiquent une partie, d'ailleurs très faible, de leur énergie cinétique. Au bout d'un nombre de collisions suffisamment élevé, l'énergie transmise est suffisante pour arracher d'autres électrons aux atomes, et le processus initial s'accélère rapidement. D'autre part, par suite du refroidissement de la chambre et de la position axiale de l'arc, un gradient considérable de température s'établit à l'intérieur de la chambre. L'ionisation sera donc faible vers l'extérieur et très élevée au contraire dans la partie axiale, où sera concentrée toute l'énergie des collisions. La température de cette zone va s'élever rapidement, en accroissant encore le taux d'ionisation. Il en résulte un premier effet de « pincement » du plasma.

Lorsque le pincement sera suffisant, l'effet électromagnétique va intervenir : les électrons en mouvement peuvent être considérés comme des conducteurs parallèles et de même sens. Ils vont s'attirer suivant l'effet Faraday, et il en résultera une nouvelle striction du plasma. On obtiendra finalement, dans la partie centrale de la chambre, un milieu à très haute température et à pression élevée par suite de l'accroissement du nombre de particules. Sous l'effet de cette pression, le plasma sera éjecté hors de la chambre par l'orifice de l'anode et communiquera son énergie au matériau placé au-devant.

Mais l'on n'obtiendra ainsi qu'une partie de la chaleur transmise. En effet, à la sortie de la chambre, les ions et les électrons vont se recombiner suivant une réaction exothermique. En ce

sens, et en ce sens seulement, le chalumeau à plasma ne diffère guère dans son principe du chalumeau à hydrogène atomique :

dans l'arc, $H_2 \rightarrow H + H$;
hors de l'arc, $H + H \rightarrow H_2$.

Le principe de l'appareil avait été établi dès 1920. Mais il fallut attendre plus de trente ans pour voir apparaître les premiers chalumeaux. Nous allons suivre leur évolution au cours des dernières années. Cette évolution a consisté en réalité en une simplification continuelle de leurs structures et de leur fonctionnement, et il est tout à fait curieux de comparer aujourd'hui la complexité des premiers modèles avec la simplicité des modèles actuels.

Le premier chalumeau fut construit en 1951. Dans une chambre en matière isolante se précipitait un violent courant d'eau. Par suite de la position tangentielle de la conduite d'entrée, l'eau était animée d'un rapide mouvement de rotation. Au cours de ce mouvement, elle délimitait autour de l'axe de la chambre une sorte de conduit cylindrique. La chambre était fermée aux deux extrémités par des électrodes de graphite et l'arc jaillissait dans le canal central. La puissance dissipée était énorme. Il est vrai qu'elle servait à la fois à évaporer l'eau, à la dissocier et à l'ioniser pour former le plasma. L'eau avait le double rôle d'élément plasmagène et de refroidisseur. La striction de l'arc était très bonne, mais l'atmosphère oxydante conduisait à une usure rapide des électrodes et à une contamination des produits traités. Les températures atteintes variaient entre 30 000 et 50 000 °K. Ces premiers chalumeaux à eau n'ont reçu que peu d'applications.

Les développements pratiques sont apparus avec les chalumeaux à gaz. Le fluide plasmagène ne participe plus au refroidissement des parois de la chambre. Il peut être introduit soit directement dans la chambre, tangentiellement à l'arc (appareil à tourbillon gazeux), soit dans une chambre auxiliaire placée en arrière de la chambre principale, à partir de laquelle il s'écoule le long de la cathode (appareil à gaine gazeuse). Dans les deux cas, il y a formation d'une couche limite froide enveloppant le plasma et protégeant les parois de l'électrode frontale.

Le choix du fluide plasmagène dépend naturellement en premier lieu du but à atteindre. Si l'on envisage le chauffage de matériaux sensibles à l'oxydation (métaux réfractaires, carbures, nitrures) ou si l'on désire assurer un fonctionnement de longue durée des électrodes, on doit naturellement éviter la présence d'oxygène. Pour des essais de laboratoire, il est intéressant d'utiliser l'hélium ou l'argon, dont le comportement est plus simple que celui des gaz diatomiques. Pour les applications industrielles, où le coût de l'opération est primordial, on utilise des gaz peu coûteux : hydrogène ou azote. Dans les essais aérodynamiques, il est intéressant d'employer l'air ou des mélanges azote-oxygène. Enfin, dans certaines synthèses chimiques, le choix du fluide est imposé par la nature même de l'opération.

Les chalumeaux réalisés au cours des trois dernières années ont atteint un très haut degré de perfection dans le domaine de la maniabilité, de l'encombrement et de la stabilité.
Les températures atteintes sont voisines de 12 000 °K. Les plus grands chalumeaux ont des puissances pouvant atteindre 10 000 kW.

L'énergie absorbée à la source d'alimentation est transformée sensiblement de la manière suivante :

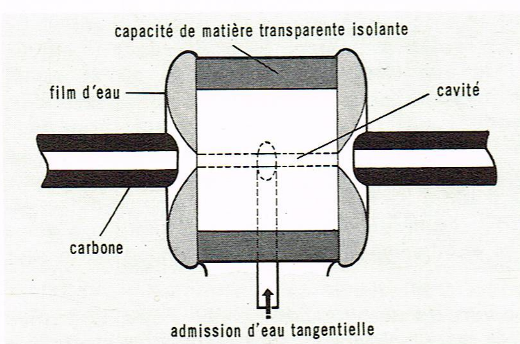
— une fraction est rayonnée sous forme d'une émission ultraviolette importante, qui peut représenter 25 à 40 p. 100 de l'énergie;

— une autre fraction (15 à 25 p. 100) est dissipée sous forme de pertes par conduction ou convexion au niveau des surfaces adjacentes au plasma;

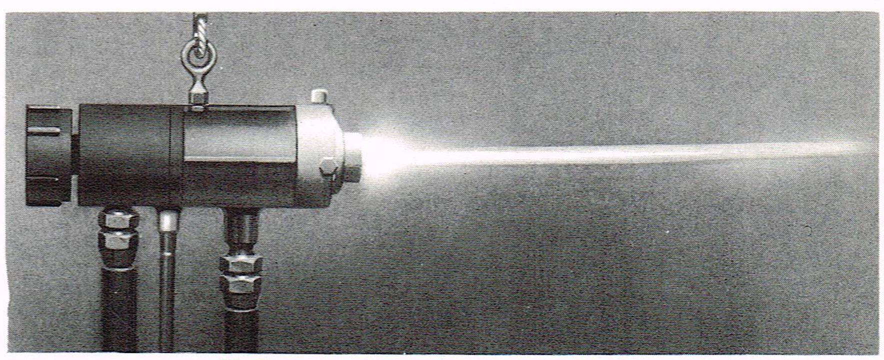
— le reste (30 à 40 p. 100) représente la part de l'énergie effectivement utilisée.

Le chalumeau à plasma est avant tout l'instrument des hautes températures associé au mouvement rapide des gaz. Nous examinerons plus loin les importantes applications qu'il a reçues dans le domaine de la recherche spatiale.

Signalons cependant, dès maintenant, que l'application de revêtements réfractaires à l'aide du chalumeau à plasma est déjà entrée dans la pratique industrielle. D'une part la



Four Maeker.



Chalumeau à plasma de 25 kW donnant une température de 28 000 °C.

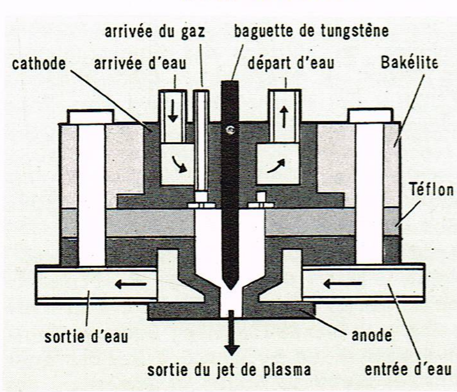
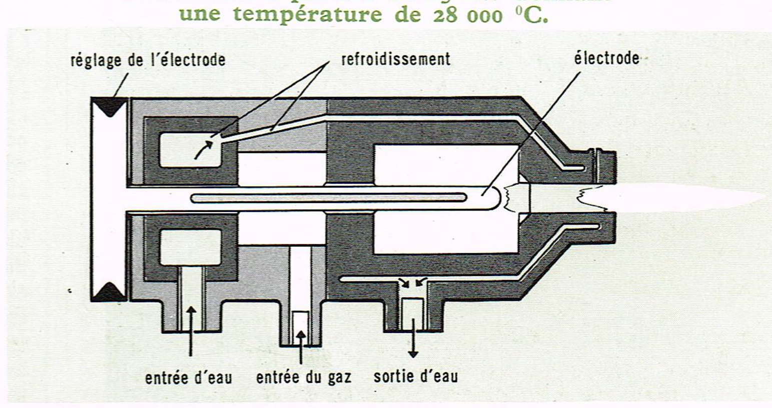
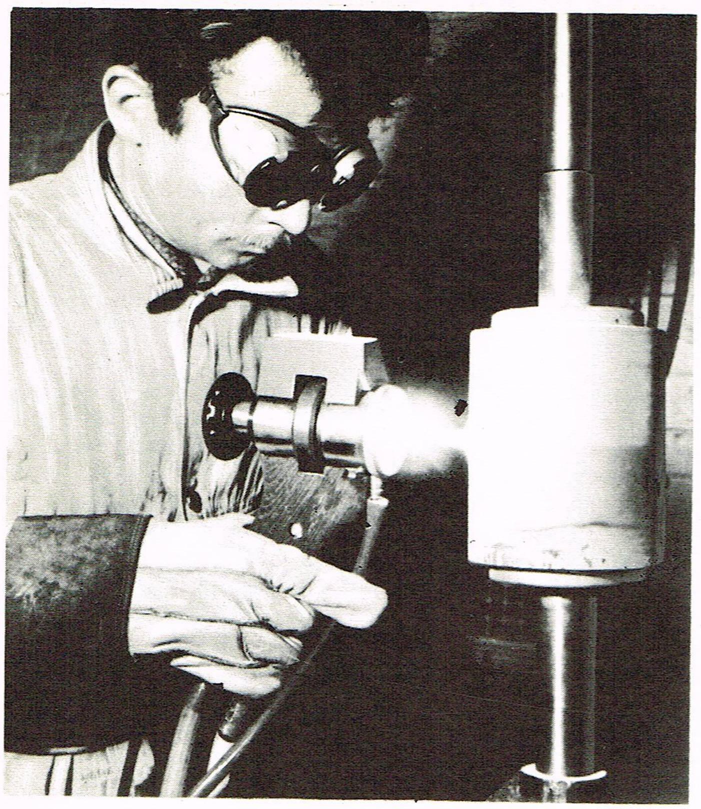


Schéma du chalumeau à plasma.

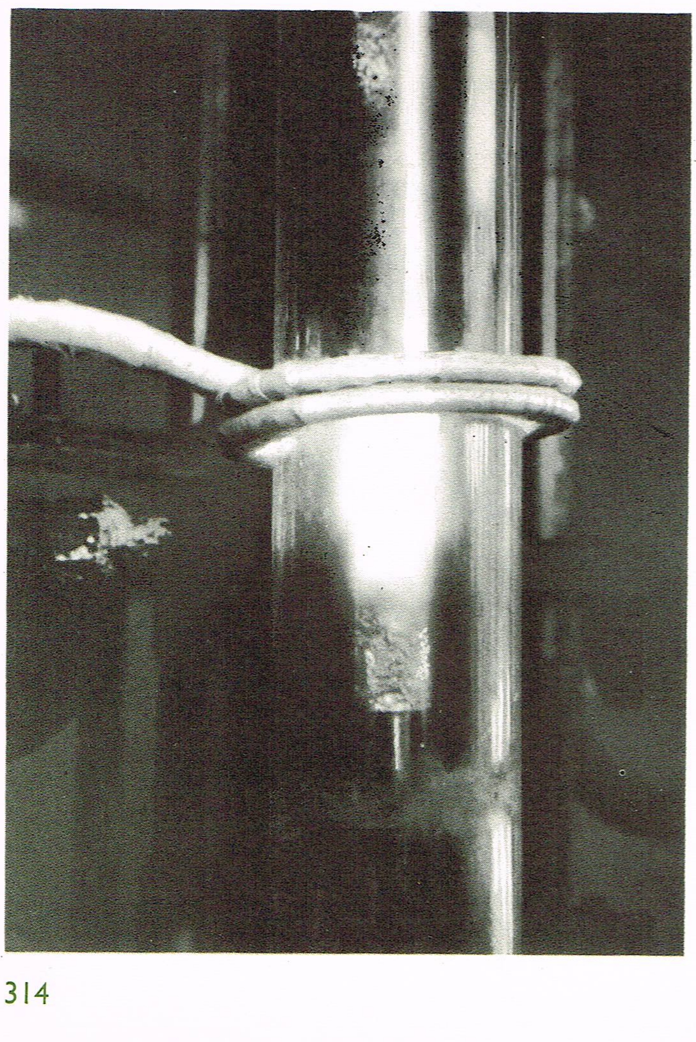


Le chalumeau à gaine gazeuse.



Le chalumeau à plasma a déjà fait son entrée à l'usine et à l'atelier pour la réalisation de revêtements réfractaires.

Purification d'un métal par zone flottante.



température de la flamme est nettement plus élevée que celle du chalumeau oxyacétylénique; d'autre part l'opération s'effectue en atmosphère inerte. Il est donc possible de déposer sur différents supports des substances extrêmement variées : des métaux (nickel, titane, chrome, molybdène, tungstène), des oxydes, des carbures, des nitrures, des borures, etc. Les revêtements obtenus sont supérieurs aux revêtements classiques du point de vue de leur adhérence, de leur compacité, de leurs résistances mécanique et chimique.

les fours à plasma

Dans cet appareil, le plasma est engendré par des courants de haute fréquence d'origine inductive. Il s'agit du même principe que dans le chauffage par induction des matériaux métalliques, mais la fréquence est adaptée au chauffage du plasma, c'est-à-dire d'un milieu à résistivité relativement élevée. Elle est donc nettement supérieure (10 à 100 fois) à la fréquence utilisée pour le chauffage des métaux.

Les principes de fonctionnement avaient été établis dès 1947. Mais, cette fois encore, l'étude fondamentale n'avait pas été suivie de réalisations pratiques à l'échelle industrielle.

C'est seulement treize ans plus tard, aux Etats-Unis en 1960 et en France en 1961, qu'apparaissent les fours à plasma à induction à haute fréquence. Leur grand intérêt provient du fait qu'il n'existe aucun contact entre le fluide plasmagène et la source de champ électrique. Celui-ci est en effet produit par un inducteur extérieur à l'enceinte de décharge.

15 000 °C dans un tube de verre !

L'élément essentiel du four à plasma est donc le générateur de haute fréquence. Le four lui-même est d'une extraordinaire simplicité. Il est constitué d'un simple tube de silice vitreuse ou d'un matériau réfractaire convenable, muni à une extrémité d'une tête assurant l'arrivée du fluide plasmagène et ouvert à l'air libre à l'autre extrémité. Pour provoquer la formation du plasma, il est tout d'abord nécessaire d'abaisser la résistivité du fluide utilisé. Ce résultat est obtenu en provoquant un effluve entre l'inducteur et une électrode de graphite ou de métal réfractaire reliée au potentiel 0. Au bout de quelques secondes, l'ionisation provoquée par cet effluve est suffisante pour permettre au courant induit de circuler à l'intérieur du fluide. Le plasma est amorcé et l'électrode auxiliaire peut être retirée. Le plasma s'entretient alors de lui-même, l'effet thermique provenant d'une part du chauffage par les courants induits, d'autre part des réactions exothermiques de recombinaison entre ions et électrons.

Les conditions optimales de fonctionnement dépendent d'un certain nombre de paramètres fondamentaux étroitement liés les uns aux autres. La nature du fluide plasmagène est généralement imposée par la nature des matériaux à traiter et par l'utilisation envisagée pour le four. Pour un grand nombre d'opérations métallurgiques, on utilisera des fluides neutres (argon) ou réducteurs (argon + hydrogène). Dans d'autres cas, on sera conduit à utiliser des atmosphères oxydantes. Pour certaines synthèses chimiques, enfin, le fluide plasmagène constituera l'un des réactifs. Ainsi la résistivité du milieu à chauffer est généralement imposée. Or, la transmission de puissance à travers un matériau chauffé par induction est optimale pour une fréquence de courant proportionnelle à la résistivité du matériau. La fréquence est donc aussi pratiquement imposée : elle sera relativement faible (1 MHz) pour un fluide assez conducteur (argon), moyenne (5 MHz) pour un fluide à résistivité modérée (oxygène, azote), très élevée (25 MHz) pour un fluide hautement résistant (hydrogène). La puissance à injecter dans le plasma dépend de la résistivité du milieu et également des

Phot. Sté Arcos et C. N. R. S. (Vitry).

dimensions de l'enceinte. Enfin, le débit gazeux dépend essentiellement de la puissance injectée. Si le rapport $\frac{\text{débit}}{\text{puissance}}$ est trop élevé, le fonctionnement manque de stabilité. S'il est trop faible, on aboutit rapidement à la destruction du four. En particulier, pour un certain nombre d'opérations industrielles, il est intéressant d'augmenter la puissance et de réduire le débit gazeux. Dans ce cas, il est avantageux de remplacer le simple tube de silice vitreuse, qui à l'origine constituait le four, par une enceinte à double paroi, en verre Pyrex par exemple, refroidie par circulation d'eau. On peut ainsi, à puissance égale, diminuer de moitié le débit gazeux, puisque le gaz circulant le long des parois n'avait pour objet que de refroidir l'enceinte sans participer directement à l'ionisation.

Les températures atteintes varient entre 9 000 et 15 000 °K, les températures maximales étant réalisées dans l'axe du four. Le four à plasma permet donc de réaliser le chauffage de matériaux quelconques dans des atmosphères quelconques (oxydantes, réductrices ou neutres), à des températures supérieures à 10 000 °K, et cela sans la moindre pollution, puisqu'il n'existe aucune électrode chaude au contact du fluide plasmagène. Il constitue donc le moyen de chauffage idéal pour le laboratoire de chimie des très hautes températures.

Par suite de ses immenses avantages, un tel appareil ne pouvait laisser l'industrie indifférente. Malheureusement, le coût du fonctionnement est relativement élevé, et cette considération risquait de freiner notablement le développement du four à plasma hors du laboratoire. Pour atteindre un stade d'utilisation industrielle, il était donc nécessaire d'augmenter la puissance de l'installation et de diminuer le prix de revient moyen du kilowatt utilisable.

Ce résultat a été obtenu en superposant au champ haute fréquence un champ électromagnétique continu ou alternatif, établi — par exemple — entre une électrode cylindrique en cuivre refroidi (constituant la cathode) placée à l'intérieur de l'enceinte, et une électrode annulaire, également en cuivre et refroidie, placée à la sortie de l'enceinte et constituant l'anode.

On peut ainsi décupler la puissance des premières installations, qui utilisaient uniquement le courant de haute fréquence, en suralimentant le four au moyen d'une énergie électrique relativement bon marché.

Contrairement au chalumeau à plasma, qui est l'instrument des hautes températures fortement localisées et associées au mouvement rapide des gaz, le four à plasma sera essentiellement l'instrument des hautes températures statiques. Dans le chalumeau à plasma, les vitesses d'éjection des gaz sont importantes, généralement de 200 à 500 m/s. Elles ne dépassent pas 10 m/s dans le four à plasma.

Le four à plasma sera aussi l'instrument de travail du chimiste qui recherche le chauffage à très hautes températures sans pollution. En effet, par suite de l'absence d'électrode chaude, il n'existe aucune limite à l'emploi de fluides très divers, même de fluides très réactifs comme le chlore. Enfin, alors que dans le chalumeau à plasma le matériau à chauffer est placé hors de la chambre, dans le four à plasma, le chauffage est réalisé à l'intérieur de l'enceinte, au sein même du plasma. Il en résulte un avantage sérieux pour le chauffage de substances très oxydables.

Il est particulièrement intéressant de constater que ces deux appareils, le chalumeau et le four à plasma, fondés sur des principes tout à fait analogues, s'adressent à des catégories très différentes d'utilisateurs. Ils se complètent admirablement et permettent au chimiste aussi bien qu'au spécialiste de l'astronautique d'accéder au domaine des très hautes températures.

un gros point noir : les matériaux réfractaires

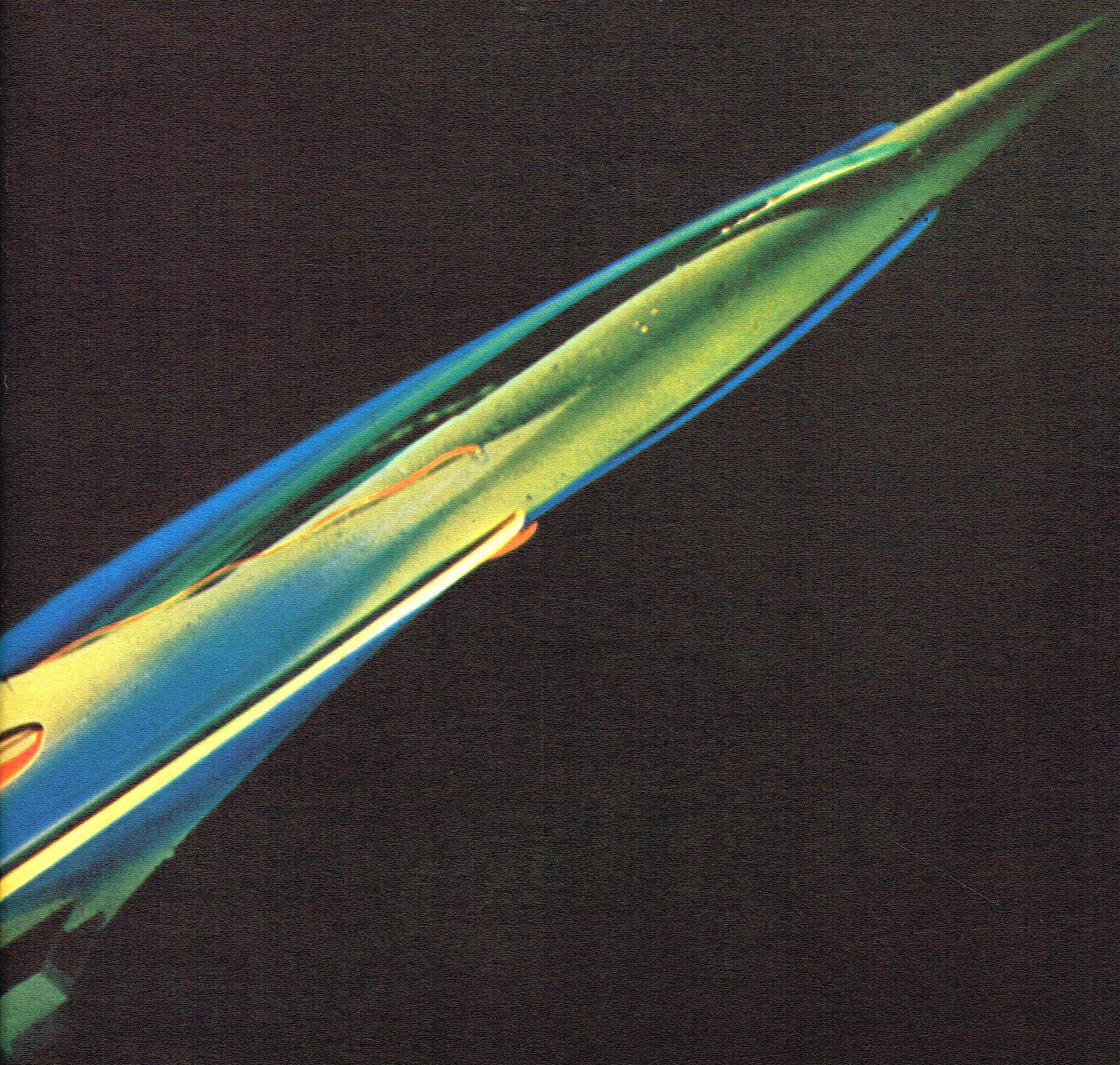
Toutes les fois que l'homme a voulu mettre en œuvre de nouvelles sources d'énergie, il s'est heurté à un problème particulièrement irritant : celui des matériaux utilisables. La grande industrie sidérurgique moderne n'a pu voir le jour que lorsque les chimistes eurent mis au point, à la fin du XIXe siècle, des briques et des creusets de matériaux réfractaires susceptibles de résister au fer fondu à des températures voisines de 1 800 °C. A mesure que la limite des températures réalisables et effectivement utilisables se déplace vers des valeurs élevées, le problème des matériaux se pose avec une acuité croissante.

Que demande-t-on à un matériau pour hautes températures? Dans la plupart des applications, les conditions imposées sont si nombreuses et si draconiennes qu'il n'existe aucun matériau présentant toutes les qualités requises. Ajoutons que dans bien des cas les conditions imposées sont contradictoires. Il faudra donc, le plus souvent, aboutir à une solution de compromis.

Si le manque de matériaux convenables pour hautes températures a été l'un des principaux obstacles au progrès, il a été aussi à l'origine d'échecs extrêmement coûteux. D'une manière générale, on peut estimer que ces échecs doivent être attribués à deux erreurs essentielles. La première réside dans l'établissement d'une hiérarchie des matériaux réfractaires fondée sur la température de fusion. Il importe peu qu'un métal ait une température de fusion supérieure à 3 000 °C s'il perd dès 1 400 °C la majeure partie de ses propriétés mécaniques. Il importe peu qu'un carbure ait une température de fusion supérieure à 3 500 °C s'il se détruit par oxydation à 800 °C en présence de traces infimes d'oxygène. Dans la plupart des cas, la température de fusion ne sera qu'un facteur secondaire. **Résistance mécanique et résistance chimique aux hautes températures seront le plus souvent les principales qualités demandées à un matériau réfractaire.**

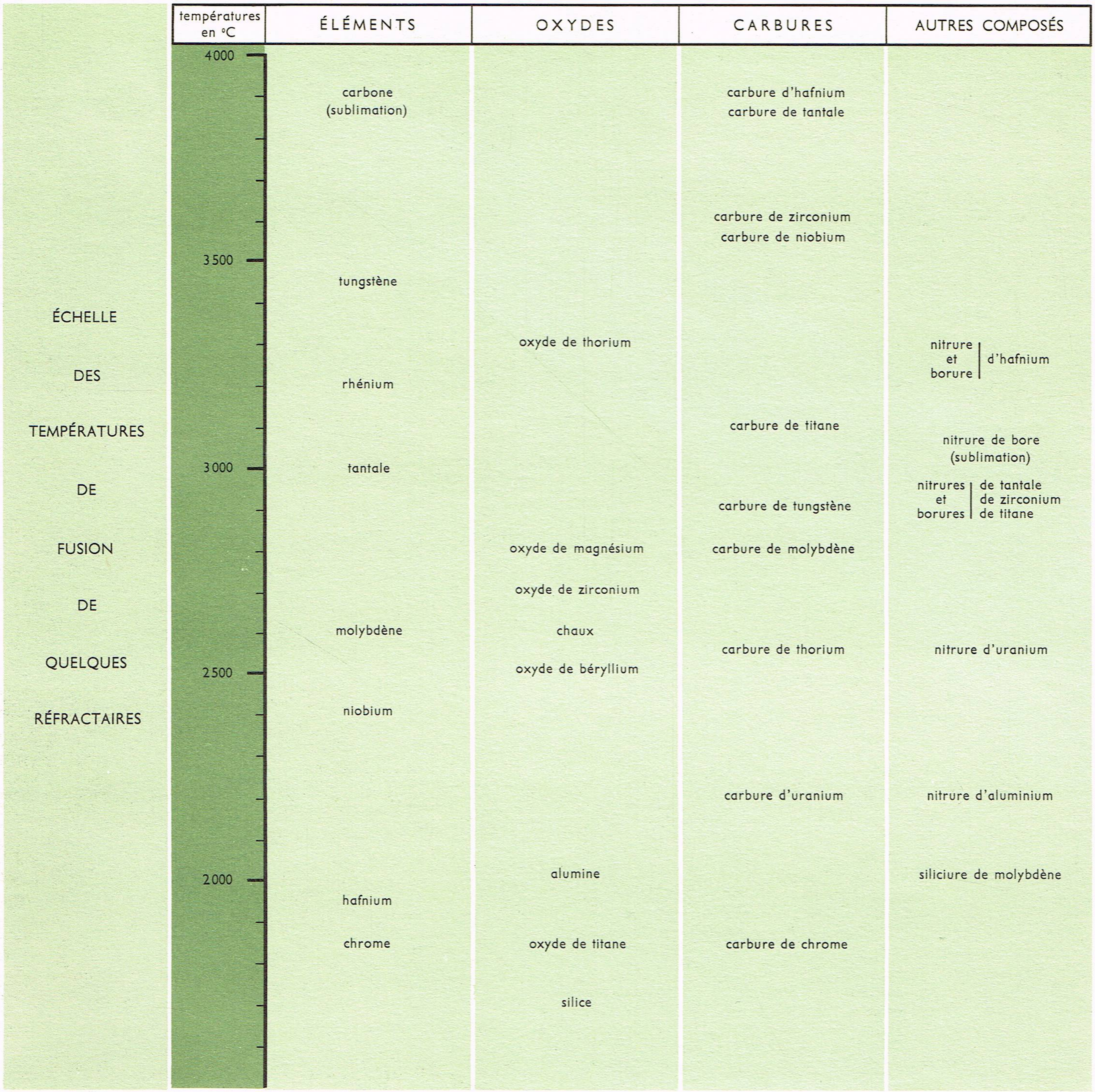
La seconde erreur, trop souvent commise, résulte de ce que l'on considère les propriétés du matériau exclusivement à la température ordinaire. Or, les propriétés aux hautes températures, c'est-à-dire dans les conditions d'utilisation du matériau, ne se déduisent pas simplement de ses propriétés à la température ordinaire. Il est certes important de savoir que le carbure de titane est à la température ordinaire l'un des plus durs de tous les corps connus, sa dureté étant 4 à 5 fois plus élevée que celle de la plupart des réfractaires. Mais il est beaucoup plus intéressant encore de savoir que cette dureté diminue très rapidement lorsque la température augmente, et qu'à partir de 800 °C elle devient inférieure à celle de réfractaires plus modestes, comme le siliciure de molybdène, à la même température. L'oxyde de béryllium est à 20 °C un extraordinaire conducteur thermique, comparable aux meilleurs conducteurs métalliques. A 1 200 °C, il est comparable du point de vue de la conductibilité thermique à la silice vitreuse, qui était 600 fois moins conductrice à la température ordinaire.





La maquette de l'avion de ligne « Concorde » est essayée ici dans un tunnel hydrodynamique, sorte de soufflerie où un liquide incolore remplace l'air. L'écoulement de l'air autour du futur avion est matérialisé par des filets de liquides diversement colorés, qui sortent de la maquette par des orifices convenablement situés.

« Concorde » doit atteindre en vingt minutes une vitesse 2,2 fois plus grande que celle du son. A ces vitesses-là, l'échauffement du revêtement posait des problèmes auxquels les techniciens ont apporté des solutions satisfaisantes.



abondance de biens ne suffit pas

Par suite des difficultés que présente la mesure des propriétés mécaniques et physiques aux températures élevées, le spécialiste des hautes températures se trouvera presque toujours placé, lorsqu'il devra choisir un matériau, devant un problème très difficile. Fort heureusement, au cours des dernières années, l'arsenal des matériaux mis à sa disposition s'est considérablement accru. Pendant longtemps, le graphite fut sa seule ressource, et il demeure encore, comme nous le verrons, son ami le plus sûr. Puis les oxydes lui apportèrent l'appui de leur incomparable inertie chimique. On peut dire que, jusqu'à la Seconde Guerre mondiale, graphite et oxydes réfractaires furent les deux principaux supports des industries mettant en jeu les hautes températures. Chacun de ces deux matériaux, oxyde réfractaire et graphite, avait son domaine d'utilisation propre, tant du point de vue des propriétés physiques que du point de vue des propriétés chimiques. Le domaine des oxydes était celui des isolants électriques ou thermiques, le domaine du graphite était celui des conducteurs. Toute opération effectuée en atmosphère oxydante faisait intervenir les oxydes, toute opération effectuée en atmosphère réductrice ou neutre faisait intervenir le graphite.

Mais, à mesure que les conditions de travail imposaient des

données plus sévères, que les structures des édifices travaillant aux hautes températures devenaient plus complexes, les oxydes et le graphite n'arrivaient plus à assurer tous les services qui leur étaient demandés. Le graphite, matériau très doux, s'use rapidement. Les oxydes sont si sensibles aux chocs thermiques que toute variation un peu rapide de la température est interdite aux appareils dans lesquels ils interviennent.

Les matériaux métalliques ont alors fait leur apparition, soulevant, il faut bien le dire, un scepticisme assez général. Or, au cours des dernières années, le succès des métaux réfractaires a été tel, que l'on commence déjà à redouter une insuffisance des ressources naturelles. Il existe dans la classification périodique des éléments une dizaine de métaux typiquement propices aux hautes températures, c'est-à-dire que leur température de fusion est supérieure à 2 000 °C. Quatre d'entre eux ont effectivement reçu des applications pour cet usage : le molybdène et le tungstène, puis, plus récemment, le niobium et le tantale. Mais la nature s'est montrée cruelle envers les métallurgistes en associant le caractère hautement réfractaire de ces métaux à une extrême sensibilité à l'oxydation. D'autre part, les métaux sont particulièrement sensibles au phénomène du fluage, c'est-à-dire à la déformation à chaud sous une contrainte donnée.

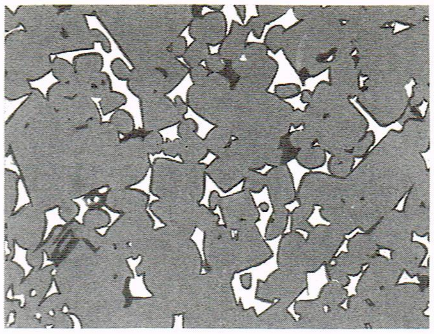
Dernier venu dans la gamme des matériaux réfractaires, le groupe des carbures, nitrures et borures comporte quelques-uns des matériaux qui possèdent les plus hautes températures de fusion actuellement connues : le carbure de hafnium fond à près de 4 000 °C. En fait, ces matériaux sont connus depuis fort longtemps, mais n'avaient pas été envisagés comme matériaux pour hautes températures.

un choix difficile

Le chercheur ou l'ingénieur qui étudie un réfractaire satisfaisant pour une application donnée se trouve finalement presque toujours placé devant un dilemme. Un matériau métallique lui assurera une bonne résistance aux chocs thermiques et mécaniques, mais il perdra ses propriétés à température moyenne et sera généralement très sensible à l'oxydation. Un matériau non métallique, notamment un oxyde, le satisfera du point de vue de la résistance mécanique à température élevée et de celui de l'inertie chimique, mais ses applications seront limitées par suite de sa fragilité aux chocs mécaniques et thermiques. Il était naturellement tentant d'associer, dans un même matériau, un métal et un composé non métallique, une céramique. Ainsi est née la classe des cermets. On a espéré, avec peut-être un excès d'optimisme, que dans cette étrange association les qualités des deux constituants se seraient ajoutées; on aurait ainsi obtenu un matériau résistant à la fois au fluage et aux chocs mécaniques, à l'oxydation et aux chocs thermiques. Des espoirs un peu démesurés semblent avoir été fondés sur l'utilisation de ces matériaux. Malgré les multiples essais, leurs possibilités réelles sont encore inconnues, car cette nouvelle classe de matériaux réunit des propriétés que l'on ne trouve associées dans aucun autre composé.

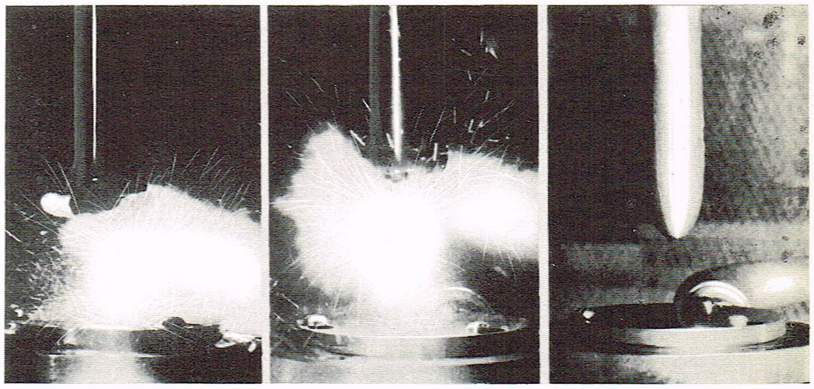
Ainsi, l'immense essor des hautes températures au cours des dernières années, avec les multiples problèmes qu'il a posés, a donné naissance à un nombre très limité de matériaux réellement nouveaux. Mais il a obligé les chercheurs à fouiller dans la gamme des matériaux existants. Certains d'entre eux, dont les possibilités d'utilisation aux hautes températures n'étaient même pas soupçonnées, se sont révélés pour certaines applications spécialisées très supérieurs aux réfractaires traditionnels.

L'un des plus difficiles problèmes posés ces dernières années aux spécialistes des hautes températures a été celui du retour des missiles dans l'atmosphère. Aucun matériau n'apparaissait susceptible de résister à l'usure intense due aux frottements des couches d'air, à l'élévation de la température et à la violence inouïe du choc thermique. Le chalumeau à plasma, en simulant

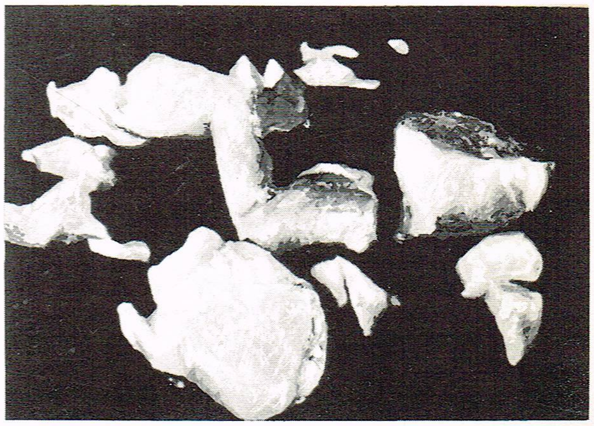


L'étrange structure du cermet : des plages de métal (nickel) enrobées dans une matrice en céramique (carbure de titane).

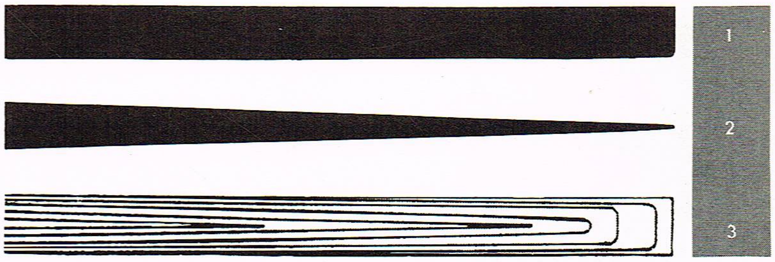
Le chalumeau à plasma juge impitoyablement les différents matériaux envisagés pour le nez de la fusée :



métal : le tungstène volatilisé en moins d'une seconde...



... oxyde réfractaire : la magnésie, que le choc thermique fait exploser...



... ou graphite, qui est érodé. *De haut en bas,* un barreau cylindrique de graphite (1), soumis à l'érosion à haute température, donne naissance à une pointe effilée (2). *En bas,* les stades successifs de l'érosion (3).

d'une manière parfaite les conditions du retour, a permis d'expérimenter au laboratoire les divers matériaux possibles. Les métaux se sont oxydés et volatilisés en quelques fractions de seconde. Les oxydes réfractaires ont volé en éclats par suite du choc thermique. Le graphite a mieux résisté, mais il a subi une érosion intense et le barreau initial a fini, en s'usant, par donner naissance à une pointe très effilée. Ce sont des super-réfractaires d'un genre vraiment très spécial qui apportèrent la solution : les matières plastiques, et notamment le Téflon. Cette solution, qui a bouleversé les habitudes de la céramique classique, n'a cependant rien de paradoxal; une chaleur spécifique élevée, une faible conductibilité thermique et le comportement du matériau devant l'ablation étant, pour cette application, infiniment plus importants que les propriétés proprement réfractaires.

Le cercle jadis très fermé des matériaux pour hautes températures s'est donc étendu fort loin, puisque les matières plastiques elles-mêmes y ont fait leur apparition. Il ne saurait naturellement être question, dans le cadre de cet ouvrage, d'envisager les avantages et les inconvénients de chacun des réfractaires, mais nous ne saurions passer sous silence les extraordinaires propriétés de celui qui, depuis un siècle, domine de très haut tous les autres : le graphite.

le graphite, matériau roi des hautes températures

Véritable matériau protée, le carbone ne cesse d'étonner les chercheurs par la diversité de ses aspects et de ses applications. A haute température et sous pression il constitue le diamant, corps d'une incomparable dureté. Dans les conditions normales, il constitue le graphite, si onctueux qu'il est utilisé comme lubrifiant.

Si le graphite est le plus vieux réfractaire connu, c'est qu'il existe à l'état naturel. Alors que la notion même de haute température était encore très vague, les propriétés réfractaires du graphite avaient déjà intrigué les chimistes. Dans un dictionnaire de chimie du XVIII[e] siècle, on relève cette définition : « On peut tenir le charbon exposé à la plus grande violence du feu et si longtemps qu'on voudra, pourvu que ce soit dans des vaisseaux exactement clos, sans qu'il éprouve la moindre altération. Les dissolvants les plus puissants de la chimie n'ont aucune prise sur lui, si ce n'est peut-être par des procédés particuliers. C'est peut-être le corps le plus réfractaire que l'on connaisse. »

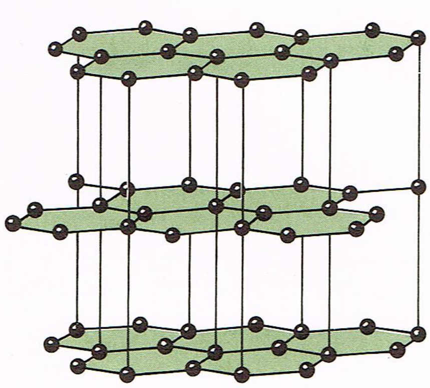
En 1808, sir Humphrey Davy découvrait l'arc électrique en l'établissant entre deux baguettes de charbon de bois au moyen d'un courant électrique fourni par une batterie de piles. L'ère des hautes températures industrielles était née, et, comme par hasard, le carbone se trouvait associé à cette découverte.

Soixante-dix ans plus tard, Berthelot précisait les relations entre les différentes variétés de carbone en distinguant deux formes cristallines : le graphite et le diamant, les autres étant considérées comme amorphes. On sait aujourd'hui que ces carbones amorphes sont en réalité imparfaitement cristallisés, avec la même structure que le graphite.

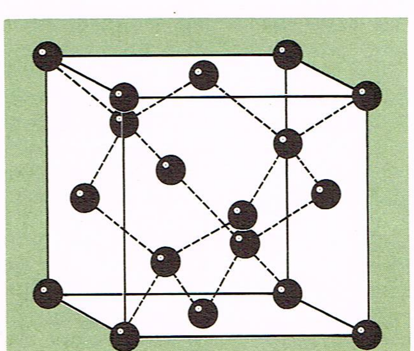
Les remarquables propriétés du carbone doivent donc être attribuées aux particularités de la structure du diamant et de celle du graphite. Le diamant a une structure cubique parfaitement symétrique, dans laquelle chaque atome de carbone est situé au centre de gravité d'un tétraèdre régulier dont les sommets sont occupés par ses quatre voisins immédiats. Le diamant est un corps isotrope. A l'intérieur de cette structure, chaque atome de carbone assure la liaison par la mise en commun de l'un de ses électrons externes avec chacun de ses quatre voisins immédiats. L'atome de carbone possédant quatre électrons externes, tous les électrons se trouvent engagés dans des liaisons et il n'existe pas d'électron libre dans la structure. Le diamant est un corps isolant. Le diamant possède une structure purement covalente. Toutes les liaisons sont équivalentes entre elles et très fortes. C'est un corps exceptionnellement dur.

La structure du graphite, déterminée par Mauguin en 1924 à l'aide de la diffraction des rayons X, est une structure lamellaire constituée de macromolécules bidimensionnelles. Les atomes de carbone se répartissent en nappes planes et parallèles liées entre elles par des forces très faibles du type de celles de Van der Waals. Le graphite est un corps tendre au clivage facile. Dans chaque nappe, les atomes de carbone sont placés aux sommets d'hexagones réguliers. Chaque atome est entouré de trois autres atomes, qui constituent ses voisins immédiats. Il est lié à chacun d'eux par la mise en commun d'un électron. Il reste donc un électron célibataire, susceptible de se déplacer et, par exemple, de s'unir avec un autre électron célibataire pour former une double liaison non localisée. Il existe donc des électrons mobiles. Le graphite est un corps conducteur. On rencontre à l'intérieur de la structure deux types de liaisons : une liaison covalente forte à l'intérieur des plans hexagonaux, des liaisons de Van der Waals faibles perpendiculairement à ces plans. Le graphite est un corps anisotrope, c'est-à-dire que ses propriétés ne sont pas les mêmes dans les différentes directions. Cette anisotropie se manifeste fortement dans toutes les propriétés mécaniques et physiques du graphite. La conductibilité électrique mesurée parallèlement aux plans hexagonaux est remarquablement élevée, comparable à celle des métaux les plus conducteurs. Elle est mille fois plus faible dans une direction perpendiculaire.

Alors que la structure du graphite est parfaitement ordonnée, celle des carbones dits « amorphes » présente de nombreux défauts de répartition et d'empilement. La longue et délicate préparation du graphite n'a d'autre objet que de réaliser la perfection de la structure.



Les structures du graphite ◀ (anisotrope) et du diamant (isotrope). ▶



Pourquoi le graphite est-il devenu le matériau type des hautes températures? Certes, il reste solide jusqu'aux environs de 3 500 °C, température au-delà de laquelle il se sublime, mais bien d'autres matériaux, le tungstène, la thorine, divers carbures, sont tout aussi réfractaires. En réalité, la propriété presque miraculeuse du graphite est celle-ci : sa résistance mécanique augmente avec la température. A 2 500 °C, alors que tous les autres réfractaires ont pratiquement perdu toute résistance mécanique, la charge à la rupture du graphite atteint une valeur double de la valeur mesurée à la température ordinaire. Aussi, bien que le graphite ait des propriétés mécaniques très moyennes à basse température, il égale les meilleurs réfractaires au-dessus de 1 200 °C et les surclasse tous à partir de 1 500 °C. Ajoutons que, son coefficient de dilatation étant très faible, le graphite résiste parfaitement aux chocs thermiques les plus durs. Il réunit donc les qualités essentielles propres aux matériaux métalliques et aux matériaux non métalliques.

Serait-il donc le matériau parfait? Jusqu'à ces dernières années, on ne lui connaissait guère qu'un défaut : sa sensibilité à l'oxydation. Tout est pourtant relatif et le graphite possède par ailleurs un ensemble de propriétés si remarquables que l'on a été pour lui très exigeant en cette matière, car si l'on met à part les oxydes, le graphite est l'un des matériaux qui résistent le mieux à l'action de l'oxygène.

Avec le développement des recherches aérodynamiques et spatiales, un autre défaut du graphite est apparu : sa faible résistance à l'érosion. Cette dernière caractéristique limite par exemple son emploi comme revêtement des tuyères de moteur de fusée, auquel pourtant ses propriétés thermiques le prédisposaient particulièrement.

Différents revêtements protecteurs ont été expérimentés, mais c'est encore le carbone lui-même qui a fourni la solution en apparaissant sous une nouvelle forme : le graphite pyrolytique.

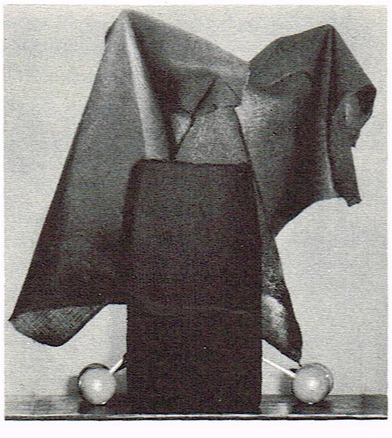
Dès 1890 Edison utilisait le cracking des hydrocarbures pour la fabrication des filaments de carbone destinés aux lampes à incandescence. Le méthane chauffé à 2 000 °C laissait un dépôt de carbone, qu'un chauffage à plus haute température transformait en graphite. La structure régulière du graphite ne se retrouve pas dans ces graphites de pyrolyse. La distance entre les plans hexagonaux n'est pas constante. D'autre part ces plans ne sont pas superposables. Conséquence de ces anomalies de structure : les graphites pyrolytiques ont des propriétés totalement différentes de celles du graphite ordinaire. Leur dureté est extrêmement élevée. Une autre caractéristique dominante de ces graphites est une très grande anisotropie. Dans son plan de clivage, le graphite pyrolytique est aussi bon conducteur que le cuivre. C'est un isolant à peu près parfait dans les autres directions. Il résiste remarquablement bien à l'oxydation et son utilisation, malgré son prix de revient élevé, est fort tentante pour la fabrication des tuyères de fusée.

Ainsi le graphite ne cesse de se présenter à nous sous les formes les plus inattendues. Mais l'une des plus curieuses est sans aucun doute constituée par le tissu de graphite. Ce tissu est réalisé par la calcination des fibres textiles les plus diverses, par exemple la rayonne, à des températures voisines de 3 000 °C. Il contient plus de 99,9 p. 100 de carbone sous forme graphitique. Cet étonnant matériau unit les propriétés physiques et chimiques du graphite aux propriétés mécaniques des textiles. D'une part, il peut être chauffé jusqu'aux plus hautes températures et voit sa résistance mécanique doubler entre la température ordinaire et 2 500 °C; d'autre part, il peut être tissé comme n'importe quel textile sous les formes les plus diverses. On envisage pour lui de multiples applications. L'une d'entre elles est déjà pratiquement réalisée : le renforcement des matières plastiques constituant les têtes de fusée.

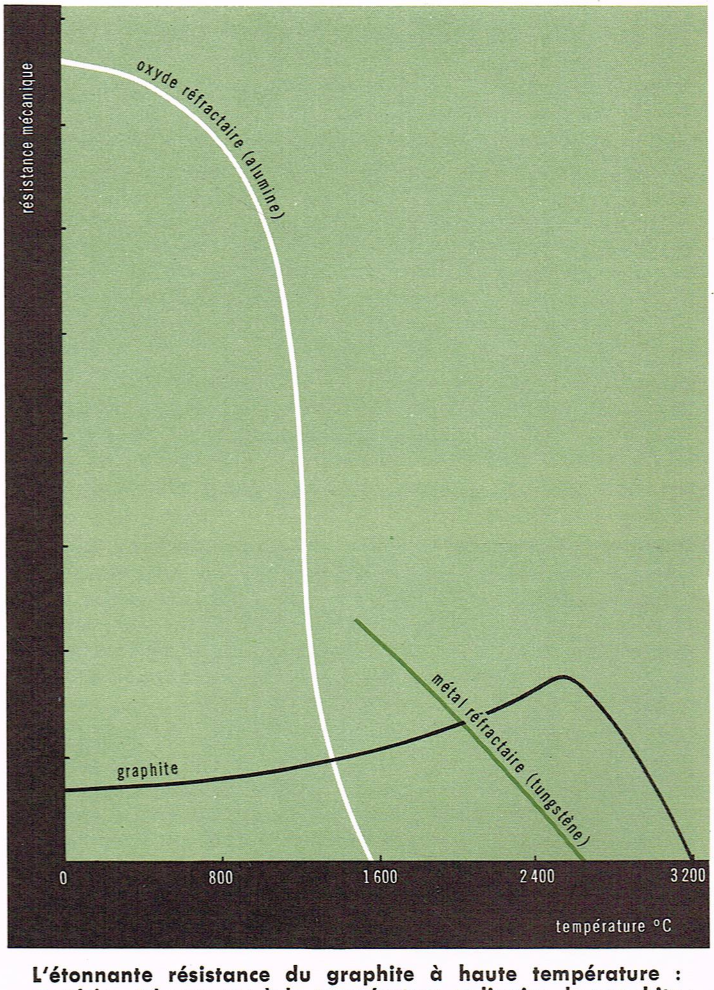
L'industrie tout entière a depuis fort longtemps fait appel au graphite en mettant à profit ses remarquables propriétés : résistance mécanique aux hautes températures, conductibilité électrique et thermique, résistance chimique. Tour à tour électrode, moule, revêtement, échangeur de chaleur, il était déjà avant la Seconde Guerre mondiale l'un des éléments essentiels des progrès des industries électriques, électrothermiques, métallurgiques et chimiques. Lorsque, en 1942, Fermi mit en route la première pile atomique, un domaine d'applications entièrement nouveau s'ouvrit pour le graphite. *Modérateur pour les neutrons, réflecteur, élément des pompes de circulation des refroidisseurs comme le sodium liquide, il est devenu le matériau fondamental de l'industrie nucléaire avant de devenir, quelques années plus tard, celui de la recherche spatiale.*



Une buse en graphite pyrolytique.



Toile en graphite : matériau des hautes températures ou tissu de haute couture?



L'étonnante résistance du graphite à haute température : matériau très moyen à la température ordinaire, le graphite surclasse tous les autres réfractaires à partir de 1 600 °C.

beaux espoirs, belles réalisations

De tous les utilisateurs des hautes températures, le métallurgiste est, sans aucun doute, le plus constant.

Pour la fabrication de son premier métal (le cuivre ou l'or?), le forgeron néolithique dut déjà faire appel à des températures supérieures à 1 000 °C. Et, si le fer fut et demeure « le cadeau que Dieu fit aux métallurgistes », ceux-ci eurent constamment besoin pour l'élaborer des cadeaux que leur firent les thermiciens. Le XII[e] siècle vit apparaître le haut fourneau; le XVIII[e], le four à creusets pour la fusion de l'acier et le four à puddler; le XIX[e] vit la naissance et le prodigieux développement des trois grandes techniques d'élaboration de l'acier fondu : le convertisseur Bessemer en 1855, le four Martin en 1863 et le four Thomas en 1878. Les débuts du XX[e] siècle furent aussi ceux des fours à arc électrique, particulièrement utilisés pour la fabrication des aciers spéciaux. Puis, au cours des vingt dernières années, les métallurgistes firent appel à toutes les méthodes nouvelles de réalisation des hautes températures, et ceci à l'échelle industrielle : four à induction, four à bombardement électronique et, tout récemment, au chalumeau à plasma pour la fusion de lingots d'acier de 150 kg. Ainsi, il n'est pas un mode de production de températures élevées qui n'ait été très rapidement appliqué par ces utilisateurs particulièrement exigeants.

des corps exceptionnellement purs

Mais si la variété et la puissance des moyens de chauffage sont l'un des aspects de l'industrie métallurgique moderne, une recherche constante de la pureté et de la constance des métaux fabriqués constitue sa caractéristique essentielle. Ce souci de perfection s'est manifesté aussi bien pour les métaux traditionnels que pour les métaux nouveaux, titane, zirconium, béryllium, etc., qui étaient encore naguère des métaux « rares ». Il s'est manifesté aussi bien au laboratoire que dans l'industrie, où l'on s'efforce d'éliminer les gaz dissous dans le métal; on a vu apparaître récemment des lingots d'acier de plus de 150 tonnes coulés sous un vide poussé.

Dans beaucoup de fabrications industrielles, les teneurs en impuretés ne se comptent plus aujourd'hui en parties ou fractions de partie pour cent; elles se comptent en parties par million et parfois même par milliard (par exemple un fer à 10^{-6} de nickel contient 1 atome de nickel pour 1 million d'atomes de fer).

Comment obtenir ces puretés exceptionnelles? Le principe consiste presque toujours à associer hautes températures et vide élevé, car les métaux sont en général très sensibles à l'attaque des atmosphères gazeuses. De nombreuses méthodes très ingénieuses, fondées sur ce principe, ont été imaginées par les métallurgistes. Mais l'une d'elles nous confond par son élégance et sa simplicité.

Tous les fondeurs savent bien qu'au cours de la solidification d'un lingot, les impuretés s'accumulent dans la partie centrale, solidifiée la dernière. Lorsque le métallographe examine au microscope un minuscule fragment de ce lingot, il constate que les impuretés n'y sont pas réparties d'une manière uniforme, mais concentrées dans les portions qui se solidifient les dernières. Son observation, effectuée sur une plage de quelques millièmes de millimètre carré, rejoint celle du fondeur, faite sur des lingots de plusieurs mètres de hauteur : au cours de la solidification, les impuretés demeurent le plus longtemps possible dans la phase liquide. Le spécialiste des diagrammes d'équilibre confirme : les éléments étrangers sont presque toujours plus solubles dans la phase liquide que dans la phase solide.

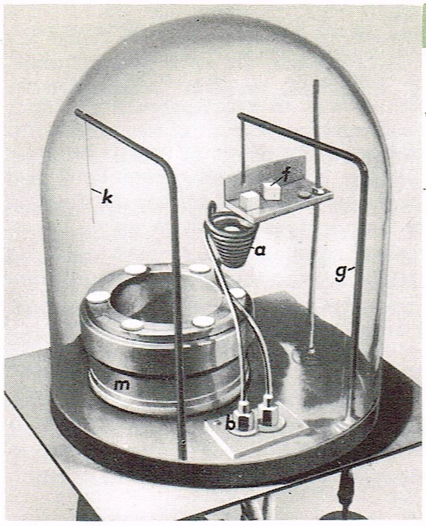
Considérons donc un barreau de métal disposé horizontalement dans une nacelle. A l'aide d'un petit four, fondons à une extrémité de ce barreau une petite portion de métal; puis déplaçons très lentement (1 mm à l'heure, par exemple) notre four vers l'autre extrémité du barreau : la partie d'abord fondue se refroidira et se solidifiera en rejetant ses impuretés vers la portion fondue voisine. Le mouvement du four continuant, chaque zone fondue puis solidifiée se purifiera en rejetant ses impuretés dans la zone suivante en fusion. Finalement, les impuretés seront drainées d'une extrémité à l'autre du barreau et se concentreront dans la zone terminale, solidifiée la dernière, qu'il suffira d'éliminer. En recommençant plusieurs fois la même opération — toujours dans le même sens naturellement —, on aboutira à une pureté exceptionnelle. A titre d'exemple, on a purifié par cette méthode un barreau d'aluminium de 35 cm de longueur et contenant 30 parties d'argent pour un million d'aluminium. Après neuf passages de la « zone fondue », la concentration en argent, sur les deux tiers de la longueur du barreau, était inférieure à 0,02 partie par million (soit un facteur de purification de 1 500); le dernier centimètre du barreau contenait au contraire 500 parties d'argent par million.

Mais pour les métaux à température de fusion très élevée, comme le tungstène et le tantale, ou les métaux très réactifs, comme le zirconium, on peut craindre une pollution par la nacelle du métal ainsi chauffé à haute température. On procède alors à l'opération de zone fondue verticale, dite « zone flottante ». Le barreau à traiter est maintenu vertical, fixé à ses extrémités dans des pinces. Le métal est fondu sur une zone étroite, la goutte liquide étant maintenue en équilibre entre les deux parties solides par les forces de capillarité. La zone flottante est déplacée de haut en bas du barreau, et l'opération de fusion par zones s'effectue sans que le métal chauffé soit en contact avec un autre matériau.

Poussant plus loin leur souci d'éviter au métal toute contamination, les métallurgistes ont imaginé l'étonnante fusion en lévitation. Le principe de cette méthode consiste à utiliser à double fin le champ électromagnétique produit par un inducteur haute fréquence de forme spécialement étudiée : d'une part pour créer un champ de forces s'opposant au champ de pesanteur, d'autre part pour chauffer et fondre l'échantillon par création de courants induits dans sa masse. Opération stupéfiante, où l'on voit le métal fondre et se maintenir en équilibre en l'absence de tout support matériel. Curiosité de laboratoire ou méthode d'avenir? Les quantités de métal fondu en une seule opération de lévitation atteignent déjà plusieurs centaines de grammes. **Et l'exemple des transistors est là pour nous prouver que des applications d'une importance capitale peuvent fort bien s'accommoder d'une production en quantité minime d'un matériau de très haute qualité.**

L'aluminium préparé par les méthodes de zone fondue titre 99,999 7 p. 100 (soit 3 atomes étrangers pour 1 million d'atomes d'aluminium). Ses propriétés mécaniques, physiques et chimiques sont sensiblement différentes de celles de l'aluminium purifié par les méthodes traditionnelles, et qui titre cependant déjà 99,99 p. 100. C'est presque un nouveau métal. Quant au germanium, il a été obtenu plus pur encore, puisqu'il contient moins d'un atome étranger pour un milliard d'atomes de germanium. On connaît ses remarquables applications en électronique : le métal ainsi purifié devient alors sensible à des additions infinitésimales judicieusement choisies.

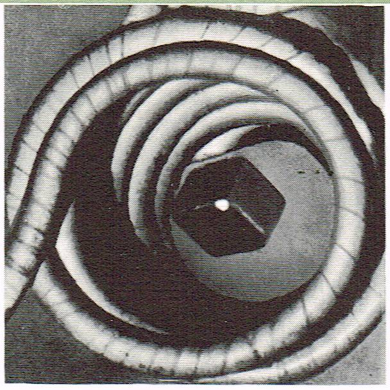
Grâce à ces méthodes de purification extrême, combinant hautes



La fusion en lévitation.

a, inducteur ; *f*, échantillons ; *g*, poussoir des échantillons ; *k*, élément d'addition sous forme de fil (pour les alliages) ; *b*, *m*, moules de coulée.

◀ L'appareillage.



▲ L'échantillon solide en équilibre.



▲ L'échantillon fondu restant en équilibre.

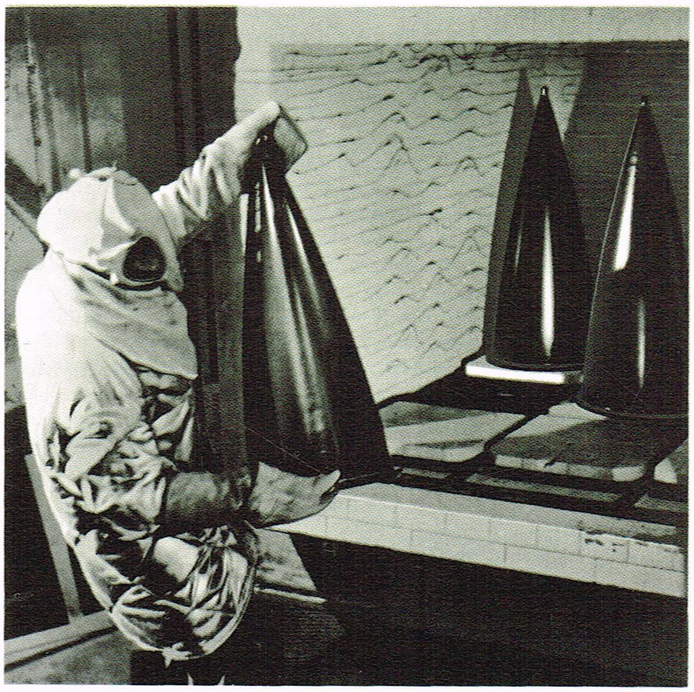
températures et très basses pressions, le métal apparaît enfin avec ses propriétés réelles, que des traces infimes d'impuretés nous avaient trop longtemps masquées.

les performances des verres

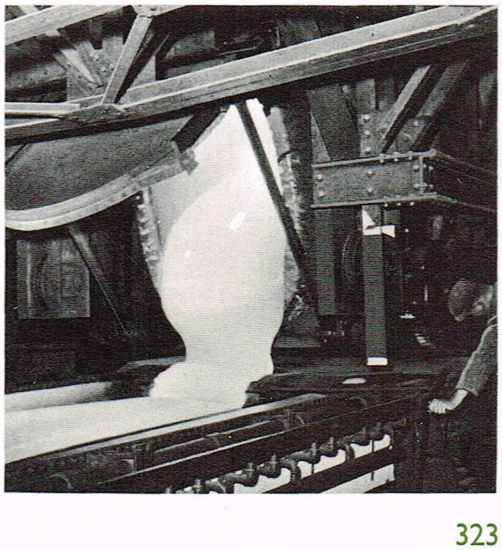
Sœur cadette de la métallurgie, la verrerie lui emprunte souvent ses modes de raisonnement et sa terminologie. Ses méthodes d'élaboration ont suivi une évolution parallèle à celles de sa grande aînée : dans les glaceries modernes les fours à pots coulés sur table ont cédé la place au véritable fleuve de glace qui s'achemine des fours vers les appareils de polissage.

Mais, par rapport au métal, le verre souffre d'une grave infériorité, qui lui a longtemps fermé les portes du domaine des hautes températures : sa sensibilité au choc thermique. Lorsqu'un matériau est chauffé ou refroidi rapidement, il s'établit des différences de température entre ses différentes parties, par exemple entre la surface et l'intérieur. Il en résulte des inégalités de dilatation ou de contraction, qui provoquent des tensions amenant rapidement la rupture de la pièce. Deux sortes de matériaux échappent à cette destruction. Les uns possèdent une conductibilité thermique élevée, assurant rapidement l'homogénéisation de la température : c'est le cas des métaux, bons conducteurs thermiques. Les autres ne se dilatent pratiquement pas par chauffage : c'est le cas de la silice, dont le coefficient de dilatation est presque nul. Aussi s'achemine-t-on vers la fabrication d'un verre de silice pure. Ce résultat est généralement obtenu par l'élimination chimique des constituants alcalins des verres ordinaires. Mais l'industrie américaine a récemment mis au point une méthode beaucoup plus audacieuse; elle consiste à évaporer la silice à haute température et à la condenser dans des fours à 1 000 °C. On a obtenu ainsi d'énormes sphères, utilisées pour la fabrication des nouvelles lentilles de télescope.

Mais là ne se borne pas l'incursion du verre dans le domaine des hautes températures, et l'une de ses dernières modifications le pose en concurrent sérieux des métaux aussi bien que des céramiques. Elle est obtenue en provoquant dans le verre la formation d'un squelette cristallin. Celui-ci prend naissance à partir de germes créés par effet photochimique (on a alors le photoceram), ou par effet thermique (c'est le pyroceram). Cette structure partiellement cristalline donne au pyroceram des propriétés exceptionnelles : il peut être à volonté opaque, translucide ou transparent, il est aussi bon isolant que les meilleures céramiques, plus léger que l'aluminium, plus dur que le silex et l'acier, et sa résistance mécanique ne commence à décroître que vers 700 °C. Transparent aux ondes de radar



Cônes de fusée en pyroceram.



Fabrication des glaces : un fleuve de verre coule sur le laminoir.

Phot. U. S. I. S. et doc. Corning Glass Work.

et résistant au choc thermique, il est peut-être le matériau idéal des cônes de fusée.

C'est encore pour l'astronautique que l'industrie verrière fait jaillir les admirables faisceaux de fibres de verre qui serviront à renforcer les matières plastiques utilisées dans les fusées.

les espoirs des chimistes

Le développement récent des méthodes de production de très hautes températures a fait naître chez le chimiste moderne de grandes espérances. Ses prédécesseurs du siècle dernier n'avaient guère disposé de températures supérieures à 2 000 °C jusqu'en 1892, année où Moissan construisait son four électrique, et portait ainsi à plus de 3 000 °C la limite des températures réalisables au laboratoire.

Il faisait aussitôt cristalliser la chaux, la strontiane, la baryte, l'alumine, le sesquioxyde de chrome. Il distillait et réduisait la silice et la zircone. Il préparait à l'état pur et en grandes quantités le bore, l'uranium, le manganèse, le chrome, le tungstène, le vanadium, le titane, le molybdène, qui n'étaient jusqu'alors connus qu'à l'état de carbures ou de carbonitrures. Il étudiait les carbures, nitrures, siliciures métalliques et métalloïdiques, jetant en particulier les fondements de l'industrie du carbure de calcium. Il distillait les métaux les plus réfractaires. Il mettait en évidence les diverses variétés de graphite, qui prennent naissance dans des conditions variables de température et de pression. Il songeait même, et probablement réussissait, à transformer le graphite en diamant. Et, cependant, le four électrique de Moissan n'avait permis de gagner que quelques centaines de degrés sur le chalumeau de Sainte-Claire Deville, qui était à ce moment-là l'instrument des plus hautes températures. Dès lors, quel progrès miraculeux ne peut-on attendre de l'introduction dans le laboratoire de chimie de techniques telles que le chalumeau à plasma, qui permet d'élever d'un seul coup de plus de 10 000 °C le niveau des températures réalisables.

Très sagement d'ailleurs, le chimiste se rend compte qu'il s'agit là d'un redoutable saut dans l'inconnu. Il sait qu'à mesure qu'augmente le niveau des températures mises à sa disposition ses repères familiers disparaissent. Jusqu'à 1 000 °C, il n'éprouve aucune difficulté à chauffer les substances dans des nacelles de quartz à l'aide de fours électriques classiques et à mesurer la température à l'aide de thermocouples. Pour atteindre 1 500 °C, il lui suffit de remplacer la nacelle de silice par une nacelle d'alumine, le four à résistance de nickel-chrome par un four à résistance de métaux précieux et le thermocouple Chromel-Alumel par un thermocouple platine-platine rhodié. Il commence déjà à éprouver de sérieuses difficultés à trouver les réfractaires convenables pour la fabrication des nacelles destinées à contenir les produits à traiter.

Mais, au-dessus de 2 000 °C, son univers se modifie totalement. Les réactions les plus lentes à la température ordinaire deviennent extrêmement rapides. Le rôle des catalyseurs et celui des réactions superficielles deviennent à peu près insignifiants. Le problème des vitesses de réaction perd de son importance et, au contraire, celui de l'équilibre devient essentiel. La chimie des solutions n'existe plus et les molécules complexes disparaissent. Seules subsistent les molécules les plus stables : l'azote, l'hydrogène, l'oxyde de carbone. Des réactions anormales commencent à se produire : l'oxyde de carbone ne brûle plus dans l'oxygène. Bien mieux, les règles habituelles de valence s'effondrent : en se combinant avec l'oxygène, l'aluminium adopte les valences 1, 2 ou 4, aussi bien que la valence habituelle 3. La thermodynamique elle-même lui réserve des surprises. Aux températures moyennes, il peut sans grand risque d'erreur confondre énergie libre et chaleur mise en jeu. Aux très hautes températures, l'entropie devient le facteur prépondérant.

Au-delà de 5 000 °C, non seulement les réactions chimiques ordinaires ont pratiquement disparu, mais les participants eux-mêmes de ces réactions, les molécules, ont cédé la place à d'autres acteurs : les atomes et les ions. De sorte que la chimie, en tant que science des molécules, pourrait bien s'arrêter au-dessous de 5 000 °C.

Même limité à 5 000 °C, le domaine à défricher reste immense. La chimie descriptive classique s'est en effet essentiellement bornée à envisager les propriétés et les possibilités de réaction des composés stables à la température ordinaire ou aux températures moyennes. Elle aura maintenant à étudier toutes les réactions homogènes et hétérogènes de tous les ions et radicaux libres stables à toutes les températures. Devant ce bouleversement de ses habitudes et de ses modes de pensée, le chimiste tente d'abord de faire le bilan de ses connaissances.

les surprises de la chimie des hautes températures

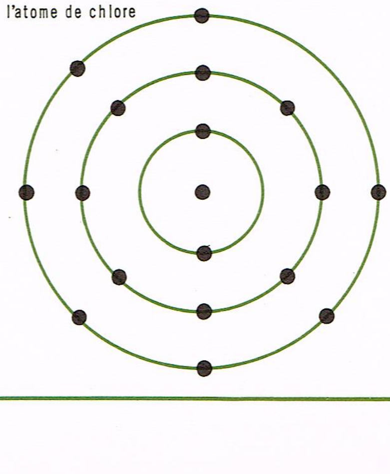
La première étape de l'exploration de ce nouveau domaine était naturellement l'identification des espèces présentes dans les systèmes de gaz en équilibre aux hautes températures. On savait qu'en élevant la température d'un système on devait nécessairement voir diminuer la concentration partielle des produits polyatomiques en faveur de structures moins élaborées : systèmes diatomiques ou même monoatomiques. Or, les chercheurs qui étudièrent par spectrométrie de masse la composition de la vapeur en équilibre avec le graphite à des températures comprises entre 2 000 et 4 000 °C eurent la surprise de rencontrer non seulement des molécules C_2 en quantité non négligeable, des molécules C_3 en proportion plus importante encore, mais également des molécules C_5 et aussi des quantités appréciables d'ions négatifs allant de C^- à C^{8-}. Bien mieux encore, la proportion des molécules C_2, C_3, C_5 augmentait quand la température augmentait.

Ce curieux phénomène est loin d'être unique. Des espèces nombreuses et parfois fort inattendues apparaissent dans la vapeur en équilibre au-dessus de solides à points de fusion élevés, notamment des oxydes ou des sels minéraux. Au-dessus d'une surface d'alumine chauffée à de hautes températures, on trouve des molécules telles que Al_2O, AlO_3, Al_2O_2. On peut aller plus loin et envisager de véritables réactions de polymérisation aux hautes températures. C'est d'une telle réaction que résulterait la formation de la suie dans la vapeur de carbone et d'hydrogène.

Des particules polyatomiques peuvent donc très bien exister à des températures élevées égales ou supérieures à 5 000 °C. C'était là une première surprise. La seconde a été plus considérable encore.

L'argon est, on le sait bien, un gaz rare de l'air doué d'une

La structure électronique de l'ion A^+ est analogue à celle de l'atome de chlore.



parfaite inertie chimique. Or, si, dans la chambre d'ionisation d'un spectromètre, on augmente progressivement la pression d'un mélange d'argon et d'hydrogène, on voit apparaître un ion AH^+ (A = argon) avec une concentration comparable à celle des autres ions A^+ et H^+. Cela signifie que la réaction d'ionisation se fait très facilement dès qu'il commence à y avoir quelques chocs en phase gazeuse. Pour surprenant qu'il apparaisse, ce résultat ne devrait cependant guère nous étonner. Il faut bien considérer que l'argon ionisé n'a nullement l'inertie chimique du gaz noble bien connu. Bien au contraire, l'ion A^+ a la structure électronique de l'atome de chlore, et l'hydrure d'argon ionisé doit avoir dans le plasma une stabilité comparable à celle de la molécule d'acide chlorhydrique dans les conditions ordinaires.

L'intervention simultanée des molécules, des atomes et des ions dans une même réaction chimique était bien propre à dérouter le chercheur. L'industrie, cependant, avait depuis plusieurs années mis en application cette chimie des hautes températures dans un certain nombre de réalisations importantes.

des réalisations... et des projets

Historiquement, c'est le problème de la fixation de l'azote atmosphérique qui a fait appel pour la première fois aux très hautes températures. Le procédé Birkeyland de synthèse du bioxyde d'azote NO par passage d'azote et d'oxygène moléculaires dans un arc électrique à courant continu est une méthode bien connue, utilisée depuis plusieurs années. La réaction globale s'écrit $N_2 + O_2 \rightarrow 2\ NO$.

Mais le mécanisme de la réaction n'a été révélé que fort récemment. La synthèse n'a lieu que s'il existe des électrons d'énergies au moins égales à 18 eV, c'est-à-dire précisément à la valeur correspondant à l'ionisation de la molécule d'azote par le processus

$$N_2 + e^- \rightarrow N_2^+ + 2e^- \text{ à } 16{,}5 \text{ eV}.$$

La réaction principale de synthèse du bioxyde d'azote serait donc

$$N_2^+ + O_2 \rightarrow (N_2O_2)^+,$$
$$(N_2O_2)^+ + e^- \rightarrow 2\ NO.$$

En fait, de nombreuses réactions secondaires interviennent, et le temps de séjour des gaz dans le plasma joue finalement un rôle fondamental. Le bioxyde d'azote formé doit être refroidi rapidement, sinon il se décompose au cours du refroidissement. Il peut être conservé indéfiniment à la température ordinaire à l'état métastable.

Ce procédé présentait de faibles rendements électriques, et il était nécessaire de procéder à une récupération coûteuse du bioxyde dilué. Aussi a-t-il été abandonné. Mais différents procédés utilisant des méthodes modernes de hautes températures, par exemple le four solaire, ont remis à l'honneur cette synthèse directe avec des résultats encore insuffisants. Les chalumeaux à plasma, eux, permettent aisément d'obtenir les 3 500 °C nécessaires à la production du bioxyde d'azote avec des rendements élevés.

Il semble, cependant, que c'est seulement dans les pays où l'énergie électrique est fournie à un prix suffisamment bas que ce procédé pourrait concurrencer sérieusement la méthode traditionnelle. Remarquons au passage combien est illogique cette méthode, qui commence à réduire l'azote à l'état d'ammoniac afin de l'oxyder ensuite.

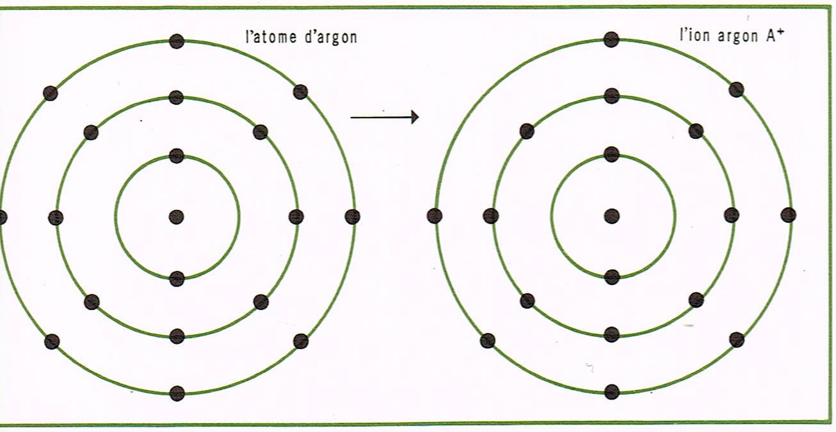
De tous les composés endothermiques dont on peut envisager la formation à hautes températures, l'acétylène est le plus important. Le chalumeau à plasma offrait aux chercheurs de nombreuses solutions possibles : alimentation par du carbone pulvérulent dans le plasma d'hydrogène, alimentation d'un chalumeau à plasma par du méthane pur, injection de méthane dans la flamme d'un chalumeau à plasma d'argon. Cette dernière méthode conduit à des rendements de l'ordre de 80 à 90 p. 100, très supérieurs aux rendements des procédés classiques de pyrolyse du méthane. Mais, seul, le chalumeau à plasma permet d'obtenir les 6 000 °C nécessaires à ces rendements optimaux.

Ne quittons pas le domaine des composés du carbone sans signaler les possibilités de synthèse du cyanogène par injection de carbone pulvérulent dans un chalumeau à plasma d'azote et de l'acide cyanhydrique dans un plasma de méthane et d'ammoniac.

Tels sont à l'heure actuelle les principaux domaines où les hautes températures interviennent dans la grande industrie chimique. Mais il est permis de penser que ces méthodes auraient un intérêt économique plus certain si elles s'adressaient à des produits de valeur marchande beaucoup plus élevée que le bioxyde d'azote, l'acétylène ou même l'acide cyanhydrique. On pense d'abord naturellement aux composés réfractaires endothermiques comme les nitrures de titane ou de magnésium, qui sont préparés en injectant la poudre métallique dans un chalumeau à plasma d'azote. Le traitement direct de certains minerais de métaux rares a même été envisagé : on a proposé d'injecter directement du minerai de béryllium dans un plasma à 12 000 °K, obtenu en faisant jaillir un arc entre des électrodes de graphite dans une atmosphère de chlore. Plus inattendue est l'intervention du chalumeau à plasma dans la synthèse des composés organiques. Des composés organométalliques très importants, comme le plomb tétraéthyle utilisé comme antidétonant dans les essences, ont été préparés en projetant dans un liquide organique un véritable brouillard de métal issu d'un jet de plasma de gaz inerte.

Dans les dernières semaines de la guerre, pour l'alimentation de leurs bombes volantes, les chimistes allemands tiraient de l'oubli un étrange composé, l'un des plus réactifs de toute la chimie, l'hydrazine, de formule $NH_2 - NH_2$. Depuis qu'Emil Fischer l'avait découvert en 1875, les difficultés de sa préparation avaient fait reculer des générations de chimistes. Or, on a réussi, à l'aide d'un chalumeau à plasma d'azote, à réaliser aujourd'hui la synthèse de ce composé, qui constitue non seulement l'un des meilleurs combustibles pour fusées, mais également le matériau de départ pour une branche entièrement nouvelle de la chimie, celle des composés azote-hydrogène, qui compte déjà plus de 500 000 représentants, aux multiples applications.

Plus problématique, mais nullement utopique, apparaît une utilisation des chalumeaux à plasma qui fut envisagée lorsque le premier « Spoutnik » fut lancé dans la stratosphère. Il s'agit de la formation, dans un plasma, de très fortes concentrations de radicaux libres ou d'atomes libres, suivie du gelage instantané de ces substances hautement réactives dans une matrice, c'est-à-dire dans un milieu refroidi à très basse température,



en l'occurrence à la température de l'hélium ou de l'hydrogène liquide. Ces milieux à haute densité d'énergie auraient pu constituer des carburants valables pour les fusées. Certes, les premiers résultats n'ont pas justifié les espérances nées dans l'enthousiasme initial. Mais cette possibilité est fort loin d'être abandonnée, et les recherches se poursuivent activement malgré les difficultés du stockage des radicaux libres.

L'application à la chimie des nouvelles méthodes de hautes températures semble donc s'effectuer d'une manière sporadique et assez timide. On peut cependant penser, en mesurant le chemin parcouru au cours des toutes dernières années, **que nous sommes à la veille d'une véritable révolution de certaines branches de l'industrie chimique.** N'oublions pas, en effet, que des appareils comme le chalumeau ou le four à plasma se trouvent aujourd'hui à un stade de développement comparable à celui où se trouvait le four électrique à l'époque d'Henri Moissan.

de la synthèse du granite à celle du diamant

Un siècle de travail acharné des physiciens et des chimistes, des chercheurs et des ingénieurs, a fourni à l'industrie de puissantes méthodes de production de températures élevées. Mais il ne faut pas oublier que la nature nous avait richement pourvus en laboratoires de chimie des hautes températures : les volcans. En général, la constitution des gaz et de la lave volcaniques correspond bien aux découvertes de la chimie des hautes températures. Mais l'origine et les caractères structuraux de certains composés trouvés dans la nature ne sont pas encore parfaitement compris. C'est que, dans la nature, les facteurs de l'équilibre sont rarement séparés. Dans toute l'histoire de notre planète, le phénomène des hautes températures a été accompagné de son inévitable corollaire : les hautes pressions.

La course vers les hautes pressions a été presque aussi mouvementée que la course vers les hautes températures. Cependant, alors que par la fusion thermonucléaire l'homme commence à savoir reproduire les conditions de température existant au cœur des étoiles, il n'a pas dépassé dans le domaine des hautes pressions les conditions qui règnent dans notre petit coin d'univers. Les plus hautes pressions réalisées au laboratoire sont de l'ordre de 2.10^5 atmosphères. Elles ne sont pas très éloignées des pressions qu'on trouverait au centre de la Terre, qui sont voisines de 10^6 atmosphères. Mais, au centre du Soleil elles atteignent déjà 10^{11} atmosphères, et 10^{15} atmosphères au cœur des étoiles blanches. Remarquons au passage que ces pressions effrayantes sont produites par l'une des forces les plus faibles existant dans l'univers : le champ de gravitation.

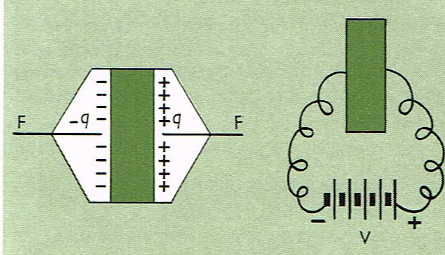
Dans leurs efforts constants pour interpréter la nature, les chercheurs ont donc été conduits à associer dans leurs expériences hautes températures et hautes pressions. Certaines tentatives dont nous retracerons l'histoire ont été couronnées d'un éclatant succès.

la synthèse du quartz

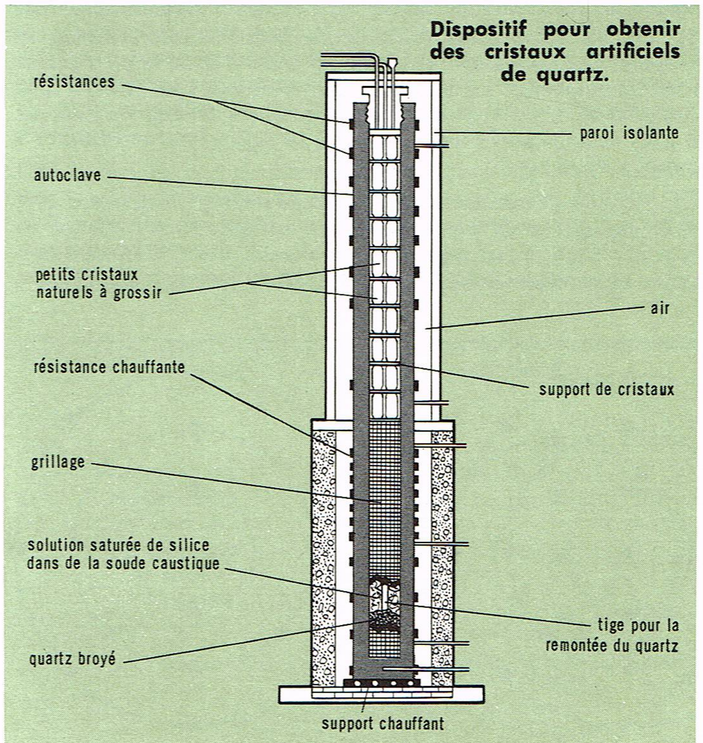
Le quartz est l'un des minéraux les plus répandus à la surface du globe, et ses magnifiques cristaux ont séduit par les nuances qu'ils peuvent prendre en présence de traces d'impuretés formées par des traces de certains oxydes métalliques (améthyste). Quand il est incolore, très pur, c'est du cristal de roche ou quartz hyalin. Mais si sa limpidité l'a fait rechercher en joaillerie, ce n'est là qu'un aspect mineur de ses applications, car il a acquis une importance considérable dans l'industrie électronique.

Considérons en effet une lame de quartz métallisée sur ses deux faces. Si l'on exerce sur cette lame une force perpendiculaire aux deux faces métallisées, il apparaît sur ces deux faces des charges $+q$ et $-q$ proportionnelles à la pression. Les signes des charges sont inversés si, au contraire, la force F est une traction. Si, inversement, on applique sur les deux faces de la lame une différence de potentiel, le quartz se contracte ou se dilate suivant le signe de cette différence de potentiel. Il peut ainsi être le siège d'oscillations mécaniques. C'est l'effet piézo-électrique.

Il est facile d'imaginer que cette piézo-électricité du quartz a reçu de multiples applications. Or, les dépôts naturels de cristaux de quartz de haute pureté et assez grands sont relativement rares, et les recherches se sont orientées vers la production de quartz synthétique. Un choix judicieux de la température et de



La piézo-électricité du quartz.



Dispositif pour obtenir des cristaux artificiels de quartz.

la pression devait permettre de reproduire artificiellement les conditions de la formation du quartz et de faire grossir de petits cristaux préexistants. Il apparut assez rapidement qu'il faut associer une pression relativement élevée à une température modérée.

L'appareil est constitué par un cylindre pouvant résister à une pression de l'ordre de 2 000 atmosphères. Les petits cristaux initiaux sont disposés dans la partie supérieure de ce cylindre, et, à la base, des déchets de quartz broyés baignent dans une solution de soude caustique. Ce dispositif est chauffé d'une manière non homogène : la solution de quartz est à 390 °C, alors que les petits cristaux à la partie supérieure sont portés à 330 °C. Par suite de cette différence de température, la solution saturée de quartz laisse déposer de la silice sur les petits cristaux, qui croissent d'une manière uniforme. La perfection des cristaux obtenus dépasse celle des cristaux naturels.

la synthèse du granite

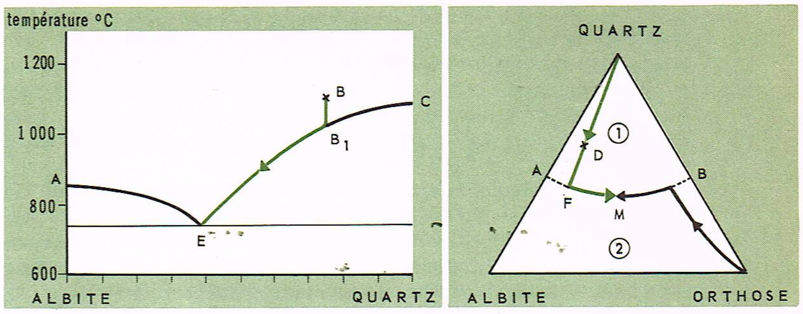
Sachant reproduire le quartz au laboratoire, les chercheurs ont pu envisager la reproduction d'une roche plus complexe, dans laquelle le quartz est l'un des constituants principaux.

L'écorce terrestre présente dans les parties continentales du globe un fort pourcentage de roches granitiques. Préoccupés depuis longtemps par la genèse du granite, les géologues ont émis de nombreuses hypothèses sur les conditions de sa formation. Selon les uns, le granite serait formé à partir de roches sédimentaires soit par action chimique d'agents en solution, soit par échange d'ions à l'état solide. Selon les autres, le granite aurait cristallisé à partir du magma fondu au sein de la Terre. En fait, ces deux hypothèses ne sont nullement contradictoires. A partir du magma ayant la composition chimique du granite, la cristallisation a pu se faire soit directement par refroidissement lent, soit en deux stades : formation d'un verre que l'on a appelé parfois « magma granitique », puis cristallisation du milieu vitreux favorisée par l'action d'agents minéralisateurs en solution.

De toute manière, il s'agit de reproduire, comme pour le quartz, un phénomène combinant haute température et haute pression, réalisé peut-être par la nature il y a des millions d'années. L'échantillon à chauffer est placé dans un creuset de platine à l'intérieur d'un cylindre, dans lequel est admise une forte pression de vapeur d'eau pouvant atteindre 4 000 atmosphères. Le chauffage de cette enceinte est assuré par un four électrique jusqu'à 1 200 °C.

Le granite est formé de trois constituants : le quartz et deux silicates du groupe des feldspaths, l'albite, aluminosilicate de sodium, et l'orthose, aluminosilicate de potassium. La plupart des granites contiennent ces trois constituants en quantités égales.

Afin d'aborder le problème par ordre de complexité croissante, les essais de synthèse ont été conduits en deux stades. Les premiers essais furent effectués avec un mélange de deux des constituants seulement : le quartz et l'albite. Un tel mélange, sous une pression de 2 000 atmosphères, commence à fondre à 740 °C, et le produit fondu contient 62 p. 100 d'albite et 38 p. 100 de quartz. Si le mélange contient précisément les deux minéraux dans ces proportions, l'ensemble sera fondu à 740 °C. Si le mélange contient un excès de l'un ou l'autre des constituants par rapport à cette composition, la température nécessaire pour obtenir une fusion totale sera plus élevée. Le diagramme d'équilibre quartz-albite présente un point d'eutexie E (le mélange dont la composition correspond au point E possède la plus basse température de fusion de tout le système). Soit un mélange quelconque composé, par exemple, de 75 p. 100 de quartz et 25 p. 100 d'albite fondus au sein du magma (point B du diagramme). Au cours du refroidissement lent, on observera d'abord



La solidification du granite.
A gauche, **diagramme de solidification du mélange binaire albite-quartz;** ***à droite,*** **section isotherme du diagramme de solidification des 3 constituants du granite (quartz-albite-orthose).**

la formation de petits cristaux de quartz (point B_1), qui se sépareront du liquide. Le refroidissement se poursuivant, la proportion de ces cristaux augmentera et, simultanément, le liquide s'appauvrira en quartz et sa composition tendra vers celle qui correspond au point E. Lorsque la température atteindra 740 °C, le liquide, dont la composition correspondra au point E, cristallisera totalement. Notons que nous avons ainsi un moyen de vérifier l'exactitude de notre hypothèse : si une roche constituée d'albite et de quartz s'est formée de la manière envisagée, les deux minéraux doivent coexister dans cette roche dans les proportions correspondant à la composition eutectique.

Considérons maintenant le mélange des trois constituants formant le granite. Nous substituons au diagramme précédent un diagramme ternaire représenté par un triangle dont chaque sommet figure l'un des constituants pur. Un mélange de deux constituants est représenté par un point sur l'un des côtés du triangle (par exemple A représente un mélange albite + quartz). Un mélange des trois constituants est figuré par un point pris à l'intérieur du triangle. Pour une pression de 500 atmosphères, le triangle est divisé en deux régions : la première, limitée par le triangle curviligne A-B-quartz; la seconde, limitée par le quadrilatère curviligne A-B-orthose-albite. Au cours du refroidissement d'un mélange représenté par un point situé dans la première région, le quartz cristallisera le premier. Pour un mélange représenté par un point de la seconde région, ce sont les silicates qui cristalliseront les premiers au cours du refroidissement. La caractéristique essentielle de ce diagramme est que la température minimale à laquelle les trois constituants cristalliseront simultanément est située sensiblement au centre du diagramme, c'est-à-dire que le mélange eutectique contient le quartz, l'albite et l'orthose en proportions égales.

Au cours du refroidissement d'un mélange liquide représenté par un point situé dans le domaine 1, D par exemple, des cristaux de quartz apparaissent les premiers et se séparent du liquide, dont la composition varie d'abord selon le trajet DF, puis selon le trajet FM. A la température correspondant au point M, tout le liquide se solidifie. De la même manière, pour un mélange représenté par un point se trouvant dans le domaine 2 du diagramme, on observe tout d'abord la formation d'orthose et d'albite, cependant que la composition du liquide varie jusqu'à atteindre la composition correspondant au point M. Ainsi, au terme du refroidissement il y a toujours la cristallisation d'un liquide de composition correspondant à M, soit un mélange en proportions sensiblement égales de quartz, d'albite et d'orthose.

Si le granite a bien été formé par solidification du magma, sa composition doit correspondre à cette composition eutectique. L'analyse de très nombreux échantillons de granite de provenances différentes a montré que les compositions les plus fréquentes se situaient bien au voisinage de cette composition, comme l'avait prévu le laboratoire.

Récemment, Jean Wyart a réussi à reproduire les conditions de

cristallisation du magma granitique vitreux. Ces essais ont porté sur un verre naturel, une obsidienne des îles Lipari ne présentant naturellement aucune trace de cristallisation et dont la composition correspond à celle du granite. Quelques grammes de ce verre sont placés dans un autoclave contenant de l'eau à 500 °C et sous 3 000 atmosphères. Il apparaît alors, avec le temps, d'abord de fines aiguilles de feldspath, puis du quartz. Dans cette opération l'eau joue un rôle essentiel. Si l'on utilise non plus de l'eau pure mais des solutions alcalines, l'action est plus rapide et plus profonde, et l'on peut distinguer du mica, des pyroxènes et des amphiboles. Ainsi, en partant d'un verre, on obtient en quelques jours une roche entièrement cristalline avec les constituants minéralogiques du granite. Les domaines de températures (400 à 700 °C) et de pressions (500 à 4 000 atmosphères) où l'on observe ces cristallisations sont bien ceux que l'on peut envisager du point de vue géologique pour la genèse des granites. L'utilisation combinée des méthodes des hautes températures et des hautes pressions a donc permis d'élucider partiellement l'un des mystères de la formation de notre planète.

de la synthèse des pierres précieuses à celle du diamant

Expliquer la nature et forcer ses secrets fut certes l'une des préoccupations constantes de l'homme. L'imiter dans ce qu'elle offre de plus rare fut l'espoir, bien souvent déçu, de générations d'alchimistes et de savants qui se succédèrent au cours des siècles jusqu'à l'aube du nôtre.

Les tombeaux égyptiens sont pleins de fausses pierres, et Pline l'Ancien, dans son *Histoire naturelle,* nous révèle la manière de les distinguer des vraies : « D'abord on pèse la pierre : les vraies sont plus pesantes. On apprécie ensuite le froid : les vraies sont senties plus froides dans la bouche. Puis on examine la substance même. » Et saint Thomas d'Aquin nous parle au XIIIe siècle des hommes qui, dans leur impureté, n'hésitent pas à fabriquer des gemmes artificielles.

Sept siècles ont passé, et la cave obscure n'est plus, où le faiseur de miracles se livrait jadis à ses jeux interdits. Les chercheurs du début du XXe siècle purent redoubler d'efforts sans redouter les foudres de l'Inquisition ni compromettre l'avenir de leur âme. La synthèse de la plupart des pierres précieuses est aujourd'hui réalisée, et les visiteurs qui assaillirent Bridgman lorsqu'il eut réussi la synthèse du diamant étaient certainement poussés par des mobiles qui n'avaient que de très lointains rapports avec la curiosité des alchimistes.

Dès 1885 apparurent sur le marché suisse des pierres rouges douées de toutes les propriétés du rubis et présentant seulement de nombreuses inclusions de bulles d'air et une coloration légèrement différente. Ces pierres provenaient de la fusion directe de fragments de rubis naturels dans le chalumeau oxhydrique et avaient été recolorées par du bichromate de potassium. C'était un résultat d'un intérêt contestable, mais il fut à la base de la véritable synthèse du corindon, réalisée par Verneuil vingt ans plus tard.

du cristal de Verneuil au rubis pour laser

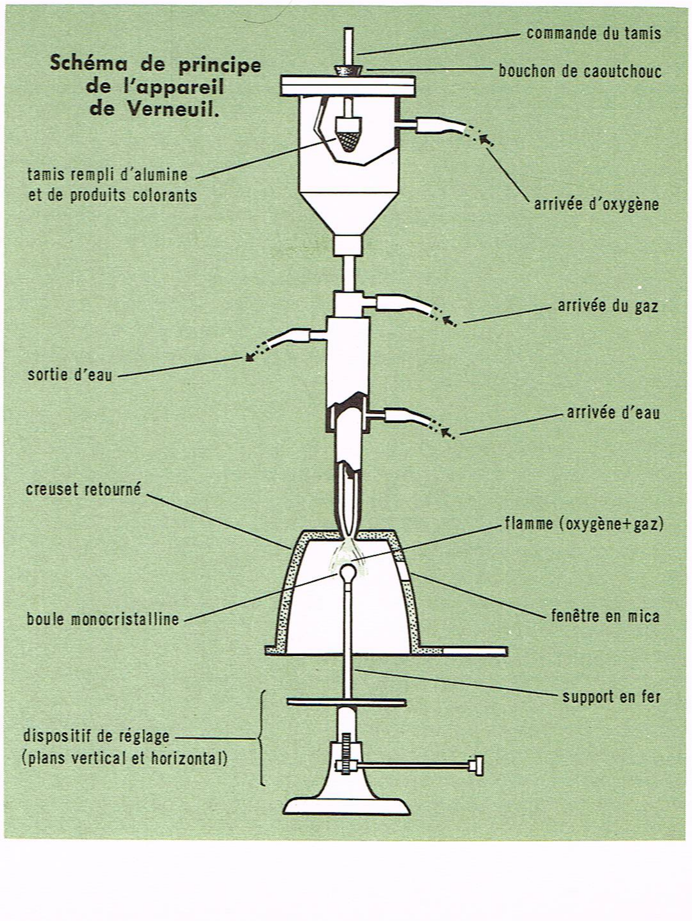
L'appareil construit par Verneuil comprenait un distributeur contenant de la poudre d'alumine très pure avec une addition de 2,5 p. 100 d'oxyde de chrome. Le distributeur recevait de légères impulsions, ce qui obligeait le mélange pulvérulent à passer par portions régulières dans une conduite aboutissant à un chalumeau oxhydrique. L'alumine fondait par passage dans la flamme du chalumeau et les gouttelettes étaient reçues sur un piédestal d'alumine frittée situé à la partie inférieure et entouré d'un creuset. Ce mode de distribution discontinue appelé « semeuse » semble essentiel pour obtenir des pierres homogènes. Par une fenêtre pratiquée dans le creuset, il était possible de contrôler la température, qui doit rester rigoureusement constante, condition indispensable à la réussite de l'opération.

Dans sa communication à l'Académie des sciences en 1904, Verneuil annonçait : « Au point de vue chimique, physique et cristallographique, il y a identité de propriétés et de structures moléculaires entre le rubis de fusion et le rubis naturel. »

Le procédé industriel est identique à celui indiqué par Verneuil. Seule amélioration apportée à la fabrication : le piédestal récepteur de l'alumine est animé d'un mouvement de rotation. Le nombre d'inclusions gazeuses est ainsi réduit, sans qu'on puisse aboutir toutefois à une élimination totale. Après refroidissement, la pierre synthétique se présente sous forme de poire. Elle est constituée d'un cristal unique, qui présente les mêmes propriétés que les cristaux naturels, propriétés optiques notamment, qui doivent être respectées lors de la taille de la pierre brute.

Aujourd'hui, les chalumeaux oxhydriques « faisant du Verneuil » s'alignent par centaines dans les usines de Suisse ou d'Allemagne, et le visiteur s'étonne devant l'incroyable démesure des bâtiments et des usines qu'il a fallu construire pour laisser filtrer au bout d'un chalumeau anonyme cette poire lumineuse, le spinelle ou le corindon. Une seule usine suisse fabrique annuellement 150 millions de carats de pierres de synthèse, ce qui représente 30 tonnes de pierres brutes.

Une des conditions essentielles pour obtenir de belles pierres transparentes est l'emploi d'une alumine de très haute pureté. L'addition de différents oxydes a permis la préparation de toute une gamme de gemmes synthétiques, rivalisant avec les pierres naturelles. L'addition de 2 à 7 p. 100 d'oxyde de chrome



donne des rubis dont la teinte rouge est plus ou moins accentuée. Le saphir bleu est obtenu par addition d'oxyde de titane et de fer.

Par addition d'oxyde de cobalt et d'oxyde de vanadium, on obtient des pierres vertes, qui ont été vendues sous le nom d'« émeraudes scientifiques ». L'émeraude n'étant pas à base de corindon, cette dénomination est éminemment frauduleuse. Ces émeraudes synthétiques ne ressemblent d'ailleurs que de très loin au béryl, silicate d'alumine et de béryllium, véritable constituant des émeraudes naturelles. Des essais de synthèse authentique à partir de fragments de béryl ont été réalisés en Allemagne, mais la fabrication délicate de ces pierres en rendait le prix de revient très élevé.

Aujourd'hui, l'industrie absorbe la quasi-totalité (90 p. 100) de la production mondiale des pierres de synthèse : industries horlogère, optique et celle des instruments de mesure, industrie textile pour les métiers à tisser résistant à l'usure causée par les nouvelles fibres. Mais toutes ces applications risquent d'être submergées par la dernière-née d'entre elles : le laser à rubis, merveille de la technique moderne susceptible de révolutionner le monde des télécommunications.

une passionnante aventure

Si, à l'aide des seules méthodes des hautes températures, la fabrication des pierres précieuses avait pu être réalisée sans difficultés exceptionnelles et atteindre rapidement le stade industriel, la plus belle des pierres n'apparaissait toujours pas dans les laboratoires. L'histoire du diamant synthétique est cependant bien longue, et les tout petits cristaux obtenus par Bridgman en 1955 sont le fruit de très nombreuses étapes.

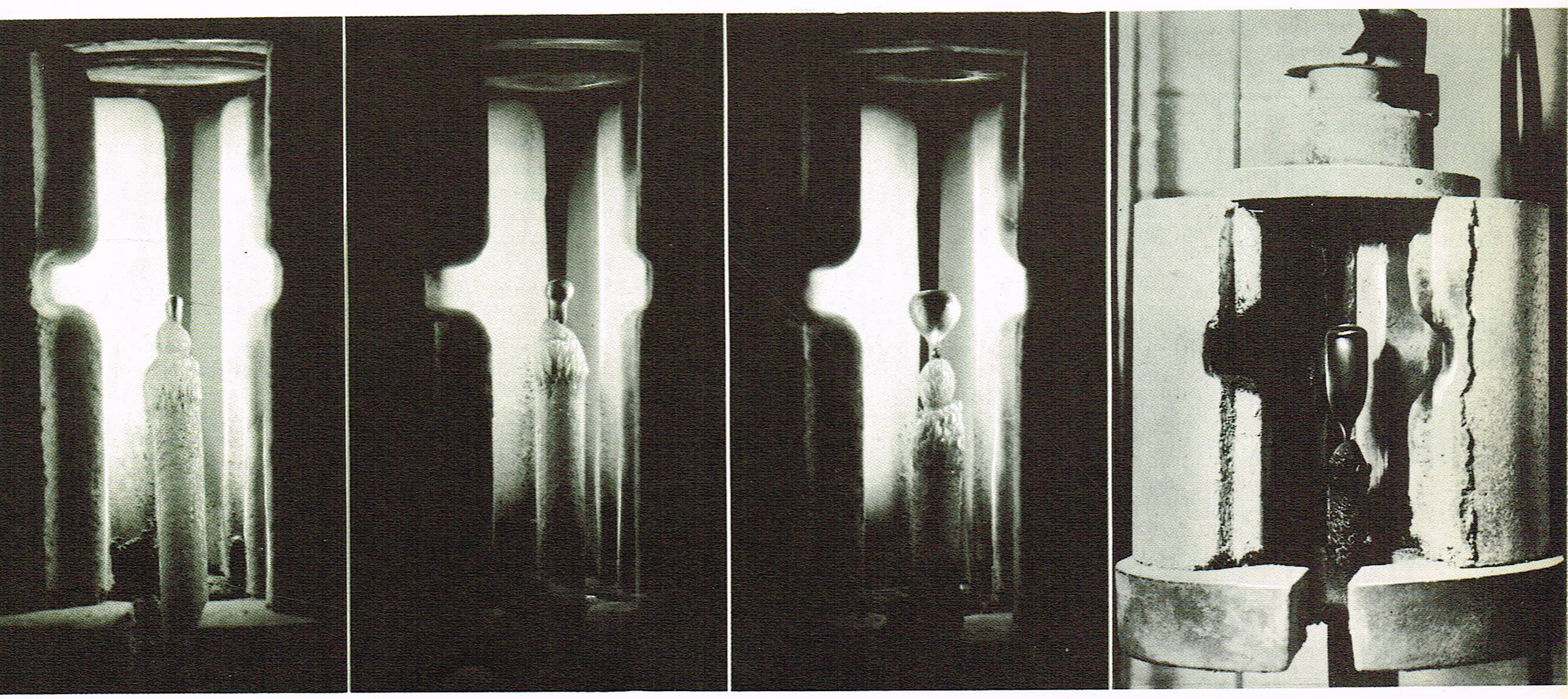
Après qu'eut été montrée l'identité de la nature chimique du diamant et du carbone amorphe, après qu'eurent échoué toutes les tentatives exclusivement thermiques de synthèse, on envisagea sérieusement les conditions de stabilité des différentes formes du carbone. Un important travail thermodynamique de Gibbs, au XIXe siècle, montre que, dans les conditions ordinaires de température et de pression, c'est le graphite qui est la forme stable. Le diamant est une forme instable dans les conditions ordinaires de température et de pression. Il est plus dense que le graphite, ce qui suggère une possibilité de le préparer à partir du carbone sous des pressions élevées.

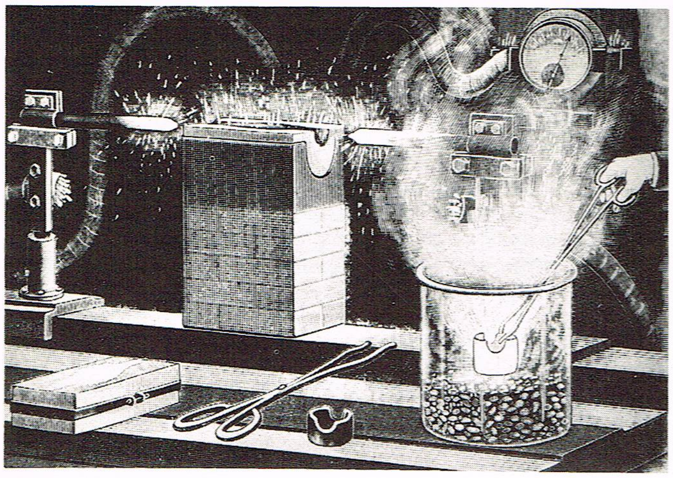
De nombreuses tentatives furent entreprises dans cette voie, la plupart n'aboutissant qu'à des échecs. En 1880, le chimiste anglais Hannay tentait de reproduire le diamant en décomposant l'huile de paraffine par chauffage dans des éprouvettes de fer scellées. On retrouve, dit-il « une masse noire, qu'on enlève au ciseau, que l'on pulvérise au mortier et dans laquelle on sent des parties très dures ». Le rendement était très faible, et ces expériences étaient de plus fort dangereuses, car de nombreux tubes explosaient. Cependant, les fragments durs et transparents recueillis présentaient les propriétés du diamant et brûlaient en donnant du gaz carbonique en bonne proportion. Toutefois, lorsque quelques années plus tard d'autres chercheurs voulurent reprendre ces expériences, ils ne réussirent pas à confirmer les résultats obtenus.

Moissan avait-il réussi ?

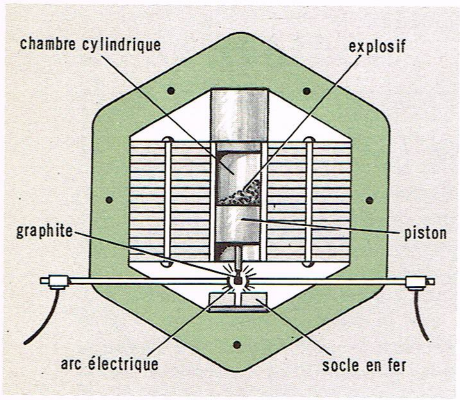
Nous parvenons ainsi en 1892, au moment où Moissan venait de mettre au point son four électrique qui permettait la réalisation de températures de l'ordre de 3 500 °C. Or, ses réflexions sur la haute densité du diamant, sur son gisement dans les roches profondes du globe, sur sa présence dans quelques météorites de fer l'avaient conduit à penser, conformément aux lois de la thermodynamique, que le diamant résulte de la lente cristallisation à haute température de la vapeur de carbone comprimée sous une très forte pression. Le problème de la température étant résolu grâce à son four électrique, restait à résoudre celui de la pression. Moissan eut l'idée d'utiliser la propriété exceptionnelle de la fonte d'augmenter de volume en

La genèse d'un rubis.

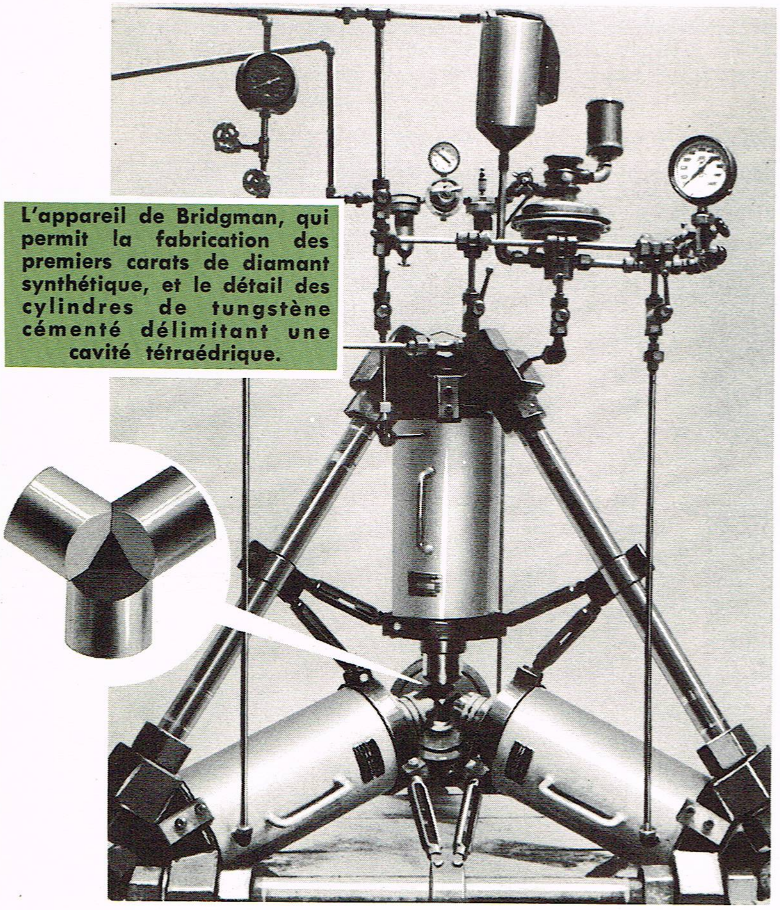




La synthèse du diamant au four de Moissan.



L'ingénieux dispositif avec lequel Majorana comptait réaliser la synthèse du diamant.



L'appareil de Bridgman, qui permit la fabrication des premiers carats de diamant synthétique, et le détail des cylindres de tungstène cémenté délimitant une cavité tétraédrique.

se solidifiant. « On chauffe pendant trois à six minutes au four électrique un creuset de charbon contenant 200 g de fer de Suède recouvert de charbon de sucre. On arrête le courant et on lève le couvercle du four. On saisit alors rapidement le creuset, on le plonge brusquement dans l'eau froide. » En se refroidissant aussitôt à sa surface, le fer forme une enveloppe solide d'acier d'une résistance considérable, qui emprisonne le reste du métal encore en fusion. Par suite de la dilatation de la fonte, il règne à l'intérieur du bloc métallique une pression énorme. Elle agit sur la vapeur de carbone qui, dès lors, cristallise à mesure que baisse la température du bain. Après attaque acide, le résidu obtenu par Moissan était constitué par du graphite et une très faible quantité de carbone plus dense, dont on pouvait séparer quelques petits diamants transparents. Moissan faisait connaître ses résultats à l'Académie des sciences le 6 février 1893.

Les discussions sur les résultats commencèrent alors. La seule méthode sûre pour contrôler la nature des cristaux obtenus, c'est-à-dire l'analyse par les rayons X, n'était pas encore connue. Il semble bien, cependant, que les expériences de Moissan aient été reproduites avec succès en Allemagne en 1917, puis aux Etats-Unis en 1924. D'autre part, la publication de Moissan fit naître une foule de vocations de « spécialistes » des hautes pressions, et de très nombreux procédés, la plupart hautement fantaisistes, furent préconisés. Retenons cependant l'ingénieux dispositif avec lequel Majorana, contemporain de Moissan, espérait réaliser la synthèse des précieux cristaux. Dans une chambre cylindrique, un piston pouvait transmettre une forte pression produite par la déflagration d'une petite quantité d'explosif. Un bloc de graphite préalablement chauffé dans un arc électrique était ainsi violemment poussé dans une cavité ménagée dans un socle de fer. Sous l'effet de la pression et de la température le graphite devait se transformer en diamant.

Le pas décisif fut franchi lorsqu'on sut réaliser des pressions de l'ordre de 30 000 atmosphères à une température de 3 000 °C, conditions qui furent obtenues à l'université Harvard durant la Seconde Guerre mondiale. L'échantillon à presser était placé entre deux cylindres de carbure de tungstène effilés et tronqués. A la température ordinaire, la pression réalisable était de 200 000 atmosphères, mais le chauffage entraînait malheureusement la détérioration rapide de l'appareil.

pas à pas vers le succès

Cependant, à chaque tentative, le problème de la synthèse du diamant approchait de sa solution. Il était apparu que la conformation des échantillons à presser jouait un rôle très important. En effet, la pression doit être transmise d'une manière uniforme. D'autre part, de nombreuses recherches avaient prouvé que les matériaux solides pouvaient être utilisés pour transmettre les hautes pressions. Ces résultats suggérèrent l'utilisation d'un solide intermédiaire entre les pistons et le corps à presser.

Le dispositif qui devait conduire au succès fut alors construit par les physiciens de la General Electric. L'appareil comprend quatre cylindres de tungstène cémenté, taillés à leur extrémité en trois faces planes et tronquées de manière à obtenir une surface triangulaire. Ces cylindres sont disposés symétriquement autour d'un point central et commandés par une presse hydraulique. Mis en contact, ils délimitent un espace tétraédrique. La conformation du volume subissant la compression est ainsi parfaitement définie. Le solide assurant la transmission de la pression est la pyrophyllite, minéral dont les propriétés physiques sont analogues à celles du talc. C'est aussi un excellent isolant électrique et thermique. La chambre de compression est constituée par un tétraèdre de cette pyrophyllite.

L'échantillon à presser est placé dans un tube qui sert également de résistance chauffante. Ce porte-échantillon est logé dans une

cavité creusée à l'intérieur du tétraèdre de pyrophyllite, cavité dont les orifices sont fermés par des pattes métalliques assurant l'arrivée du courant électrique.

A l'aide de ce dispositif, des pressions de l'ordre de 30 000 atmosphères, associées à des températures de 3 000 °C, ont pu être réalisées. Elles ont permis la fabrication de milliers de carats de diamant synthétique.

Le prix de revient est actuellement plus élevé que celui des diamants naturels, mais les propriétés des cristaux obtenus sont légèrement différentes. Ainsi, le pouvoir abrasif est supérieur. Des préoccupations d'ordre économique vont accélérer les recherches et la production. Dès maintenant, l'industrie absorbe 80 p. 100 de la production mondiale pour la fabrication de meules, de scies diamantées, de filières. L'industrie pétrolière, en particulier, fait une grande consommation d'outils diamantés pour son matériel de forage.

plus dur que le diamant

L'étude comparée des structures du graphite et du diamant nous a permis de comprendre les propriétés remarquables et si différentes de ces deux formes d'un même élément : le carbone. Mais la chimie structurale n'a pas pour objet unique l'interprétation de phénomènes connus; son rôle est aussi de susciter les recherches permettant la synthèse de composés nouveaux.

Il y a environ un siècle, Henri Sainte-Claire Deville préparait par action de l'ammoniac sur le bore un composé très difficile à fondre, peu sensible à l'action de l'oxygène et d'apparence amorphe : le nitrure de bore. Plus tard, des études par rayons X montrèrent que ce corps était en réalité cristallisé : les atomes de bore et d'azote alternent aux sommets d'hexagones réguliers disposés dans des plans parallèles et équidistants. C'est précisément la structure du graphite.

Allait-on alors, par un traitement convenable, obtenir ce nitrure de bore sous une forme identique au diamant? C'est bien ce qui a été constaté : la mise en œuvre de moyens analogues à ceux utilisés par Bridgman (pressions supérieures à 10 000 atmosphères à 1 700 °C) a permis d'obtenir des cristaux cubiques du type du diamant, dont les dimensions étaient de l'ordre du millimètre. Ils seraient insensibles à l'action de l'oxygène à 1 900 °C et leur dureté serait supérieure à celle du diamant. La préparation de cette nouvelle variété du nitrure de bore, le borazon, constitue l'une des plus belles réussites de la chimie minérale moderne.

La fréquence des colloques internationaux qui réunissent chercheurs et ingénieurs spécialistes des hautes températures ne cesse d'augmenter. Les participants éprouvent chaque fois la sensation de se trouver en présence d'un univers nouveau, tant est riche chaque année la moisson des résultats inattendus ou longtemps espérés. Ils ressentent aussi l'impression — bien moins agréable — de voir s'accroître la difficulté des problèmes scientifiques ou techniques qui leur sont posés. N'est-ce pas l'un des critères de la réussite?

Critère de réussite également l'accroissement du nombre des « solliciteurs ». Nous l'avons dit, le métallurgiste fut l'un des premiers. Mais combien de recherches d'avant-garde attendent aujourd'hui du physicien ou du chimiste des hautes températures la découverte de la méthode ou du matériau qui leur permettra de progresser : recherche spatiale, énergie nucléaire, télécommunications, conversion magnétohydrodynamique de l'énergie, etc.? Un seul exemple : le laser, la plus belle peut-être des découvertes contemporaines, qui ouvre à tous, de l'astronome au biologiste, d'éblouissantes perspectives, comporte comme pièce maîtresse le cristal de rubis, pur produit de la technique des hautes températures.

A l'observateur superficiel, l'histoire des hautes températures apparaît singulièrement féconde en rebondissements et en retours imprévisibles. Elle est, cependant, dominée par un phénomène essentiel qui n'a cessé de se préciser au cours des dernières années : la réduction vertigineuse du laps de temps qui sépare la découverte au laboratoire d'une méthode ou d'un matériau nouveau de son application industrielle à une échelle souvent gigantesque. Rien n'indique mieux à quel point l'industrie moderne tout entière est tributaire des progrès réalisés dans le domaine des hautes températures et à quel point hommes de laboratoire et réalisateurs sont solidaires dans leur œuvre.

Combien apparaît aujourd'hui dérisoire et périmée la traditionnelle distinction entre science pure et science appliquée!



**La panoplie du parfait utilisateur américain des résines synthétiques vers 1950 comprend déjà la chemise, le pantalon et la ceinture du personnage, plantureux et visiblement satisfait.
Il suffit de se livrer chez soi à un inventaire tant soit peu détaillé, portant sur l'habillement, les appareils électroménagers, les produits d'entretien, la voiture, la cuisine, le mobilier, le jardin, pour découvrir au moins autant d'objets en « plastique », qui font désormais partie de notre vie de tous les jours.**

Phot. U. S. I. S.

LES MACROMOLÉCULES

DES MATIÈRES PLASTIQUES A L'A.D.N.

par

Jacques PARROD

Le terme de « macromolécule » porte en lui sa signification; il s'agit de molécules géantes, le gigantisme commençant conventionnellement à des ensembles de quelques milliers d'atomes, sans qu'il y ait lieu de lui fixer une limite supérieure.

C'est là un domaine immense, dont l'exploration systématique a commencé depuis quelques dizaines d'années seulement. Son développement était en effet conditionné par la connaissance préalable des substances cristallisables ou distillables, et par la découverte de techniques physico-chimiques de plus en plus fines. La notion même de macromolécule ne s'est dégagée que lentement. Discutée, et même combattue, elle fut défendue par Staudinger dès 1920. En France, c'est à Georges Champetier que revient le mérite d'en avoir été le pionnier.

Les sciences macromoléculaires connaissent un essor d'une ampleur comparable à celui des sciences nucléaires, moins éblouissant, mais peut-être plus ambitieux encore. C'est en effet dans la finesse des détails de la structure de certaines macromolécules que l'on croit pouvoir trouver maintenant l'image de tout être vivant, et l'ultime raison de son développement, c'est-à-dire le secret de la vie.

Moins nobles, mais très appréciés sont les bienfaits matériels que nous dispensent quotidiennement les macromolécules; en quelque lieu qu'on se trouve, le regard rencontre des objets fabriqués avec des composés macromoléculaires synthétiques. Ils rivalisent avec des produits que la nature fournit; la soie naturelle doit lutter contre des dérivés immédiats de la cellulose, comme les acétates, susceptibles d'être filés. Ces derniers à leur tour subissent la concurrence de produits tels que le Nylon, le Rilsan et autres polyamides, plus proches chimiquement de la soie naturelle, mais entièrement synthétiques. Le caoutchouc naturel, précieux constituant du latex de l'hévéa, ne survivrait pas à une baisse importante du prix de son monomère, l'isoprène, et la dure expérience des années de guerre a démontré qu'il n'est même plus le substrat irremplaçable des pneus et de quantité d'objets. Le bois, le cuir voient leur empire se rétrécir; il n'est pas jusqu'aux métaux et aux matériaux de construction qui ne soient exposés à l'invasion insidieuse et tenace des matières plastiques.

Nous n'avons pas ici pour objet de procéder à une revue des composés macromoléculaires industriels et de leurs usages; sur ces questions, les documents désirables sont nombreux et facilement accessibles. Mais, comprendre ce que les différents produits macromoléculaires ont de commun, saisir à quelles sources originelles ils puisent leurs modes de préparation et leurs propriétés, percevoir les directions principales des recherches actuelles et les espérances qu'elles suscitent, voilà des tâches qui impliquent un effort laborieux. Le lecteur qui aurait la curiosité d'ouvrir un ouvrage à prétentions scientifiques traitant des macromolécules le refermerait en effet bien vite, à moins d'être rompu lui-même à la pratique de la chimie, de la physique et des mathématiques. Notre ambition est de faire apparaître sous un jour moins sévère ce qu'il y a de réellement fondamental dans ces problèmes.

A ceux qui ne possèdent de la chimie qu'une idée un peu brumeuse, il est conseillé de lire les premières pages; ils se convaincront que le chimiste voit les molécules aussi concrètement que le botaniste les feuilles des arbres. Les autres chapitres n'en paraîtront que plus aisés; ils peuvent être parcourus dans un ordre quelconque. Chacun projette quelque lumière sur un sujet particulier, mais bien souvent les notions exposées possèdent une valeur générale manifeste; tel est le cas des pages consacrées au polystyrène.

LES ENCHAINEMENTS D'ATOMES

Atomes et molécules, autrefois entités mystérieuses, à peine dégagées des nuées de l'alchimie, ont pris peu à peu poids, forme et dimensions. Ce sont des objets matériels comparables à tous les autres; chimistes et physiciens les manient à la façon des architectes. Et cela au sens propre du terme, puisque, le commerce s'en étant emparé, on peut acheter des « atomes » en bois, taillés à une échelle déterminée, qu'un dispositif ingénieux de boutons-pression permet d'enchaîner les uns aux autres, afin de construire des modèles de molécules. Les savants et les marchands ont tiré parti du fait que quelques règles fort simples président à l'organisation de la matière.

On peut, en chimie, ignorant les phénomènes de transmutation, considérer — comme au temps de Démocrite — que la matière est formée d'atomes séparés par du vide. Ces derniers, en nombre immense dans l'univers qu'ils meublent, se répartissent en une centaine de familles : hydrogène, hélium, lithium, glucinium, bore, carbone, azote, oxygène, etc. Chacune est subdivisée en quelques sous-familles d'isotopes, complication dont nous n'aurons pas à tenir compte, car tous les isotopes d'une même famille possèdent les mêmes propriétés chimiques et des dimensions identiques; puisqu'ils se rencontrent mélangés dans les mêmes proportions, il est commode d'utiliser, dans la pratique, un atome conventionnel, dont le poids est une moyenne. Le poids atomique étant celui de $N = 6{,}02.10^{23}$ particules, il suffit, pour connaître le poids d'un atome moyen, de consulter un tableau des poids atomiques et de diviser le nombre indiqué par N.

COMMENT SONT CONSTITUÉES LES MOLÉCULES LES PLUS SIMPLES

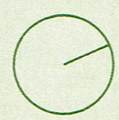
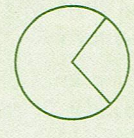
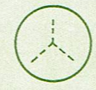
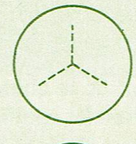
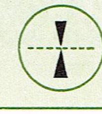
On peut donc se représenter chaque atome comme une sphère, dont le rayon (« rayon de covalence ») varie d'une famille à l'autre. On le mesure en angströms, valant un cent-millionième de centimètre (10^{-8} cm). Lorsque deux atomes se combinent, leurs sphères de covalence viennent en contact. Voilà déjà une première indication sur la structure des enchaînements d'atomes, appelés molécules.

Certains atomes, comme celui d'hydrogène, ne peuvent se lier qu'à un seul autre, tandis que l'oxygène se lie à deux atomes, l'azote à trois, etc. De façon imagée, chaque atome se comporte comme s'il tendait un certain nombre de « mains », ne demandant chacune qu'à se saisir d'une autre main, portée par un atome voisin. Ce langage trivial n'est pas, on s'en doute, celui des chimistes; ils substituent à « main » le terme technique de « valence ». Chaque valence est représentée symboliquement par un trait.

Dans le tableau ci-dessous, la dernière colonne contient l'indication du nombre de valences de chaque espèce d'atomes : l'hydrogène en porte une, et l'oxygène deux. Les figures correspondantes fournissent d'autres précisions, telles que l'angle que les valences font entre elles. Cet angle, en effet, est toujours le même pour les atomes d'une même espèce, quelles que soient les combinaisons auxquelles ils participent. Pour l'instant, nous en resterons sur ces affirmations catégoriques, nous réservant d'introduire quelques modifications par la suite.

On remarquera que, tout comme l'hydrogène, le chlore ne porte

QUELQUES ATOMES ET LEURS VALENCES

noms des corps simples	symboles	poids atomiques en grammes	rayons de covalence en angströms	modèles : 7,5 mm correspondent à 1 angström	nombre et angles de valences	
HYDROGÈNE	H	1,007 97	0,32		1	
CHLORE	Cl	35,453	0,99		1	
OXYGÈNE	O	15,999 4	0,66		2	105°
SOUFRE	S	32,064	1,04		2	105°
AZOTE	N	14,006 7	0,70		3	108°
PHOSPHORE	P	30,973 8	1,10		3	100°
CARBONE	C	12,011 15	0,77		4	109° 28′

qu'une valence; le soufre, comme l'oxygène, deux. Les trois valences de l'atome d'azote forment deux à deux un angle égal à 108° (elles ne se trouvent donc pas dans un même plan, sinon la valeur de cet angle serait de 360°/3 = 120°).

L'atome de carbone porte quatre valences, formant entre elles quatre angles égaux chacun à 109° 28′. Pour se représenter leur disposition mutuelle, il faut bien voir que le plan déterminé par deux quelconques d'entre elles est perpendiculaire au plan des deux autres.

Le lecteur peut maintenant trouver facilement la forme et les dimensions des molécules d'ammoniac NH_3, de méthane CH_4, de chloroforme $CHCl_3$, etc.

Voilà donc ces particules dont les Anciens avaient affirmé l'existence; ils ajoutaient qu'elles se mouvaient, et combien cela est exact! Non seulement les molécules se déplacent en ligne droite jusqu'à se heurter et rebondir l'une sur l'autre, mais elles tournent sur elles-mêmes et sont en proie à de trépidants mouvements de vibration internes. Les valences qui réunissent leurs atomes doivent être comparées, en effet, à des ressorts plutôt qu'à des tiges rigides.

Bien plus, les atomes tournent librement, ou peu s'en faut, autour des liaisons simples; les conséquences de cette propriété seront exploitées plus tard; elles sont importantes pour les enchaînements d'atomes de carbone, clefs du domaine macromoléculaire.

DE LONGUES CHAINES D'ATOMES

Les principes qui viennent d'être exposés ne s'opposent pas à ce qu'un premier atome de carbone tende la main à un deuxième, celui-ci à un troisième, etc. On dit alors qu'ils forment une chaîne sur laquelle de nombreuses valences restent disponibles pour d'autres atomes qui, en s'accrochant à elles, compléteront la chaîne carbonée pour former une molécule stable. Lorsque ce complément est constitué uniquement par de l'hydrogène, on a une molécule d'un hydrocarbure saturé, telle celle du pentane.

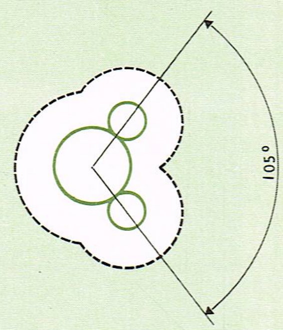
Mais pourquoi se limiter à une chaîne de 5 atomes? Rien ne s'oppose à ce qu'on lui ajoute un sixième carbone, un septième, etc. Si toutes les valences restantes sont saturées par de l'hydrogène, on a une suite de constituants du pétrole brut, du moins de celui des gisements de Pennsylvanie ou du Michigan. Ces substances (hydrocarbures saturés normaux) sont liquides à la température ordinaire à partir du pentane, solides à partir de l'hexadécane, dont la molécule a pour squelette une chaîne de 16 atomes de carbone.

Les différentes qualités de paraffines sont des mélanges des produits solides précédents, dans des proportions variables. Des molécules beaucoup plus longues contiennent 1 000, 10 000, 100 000 atomes de carbone enfilés bout à bout : ce sont les polyéthylènes, servant à fabriquer une multitude d'objets (cuvettes, seaux, etc.). Au toucher, par leur aspect, leur résistance aux agents chimiques acides ou alcalins, ils se rapprochent beaucoup des paraffines; mais ils sont plus solides et fondent à température plus élevée; celle-ci ne dépasse pas 120 °C pour les produits industriels courants, mais cela suffit pour l'usage que l'on en fait.

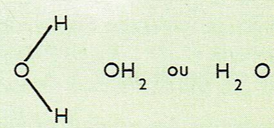
Nous voici tout simplement en plein dans le domaine macromoléculaire, le préfixe « macro » signifiant que le nombre des atomes qui constitue une molécule est élevé, supérieur à 1 000, par exemple.

Le jeu de construction des molécules de la chimie organique se poursuit sans peine. D'autres atomes que l'hydrogène peuvent se lier au carbone, ainsi qu'il apparaît dans la formule de la molécule d'alcool amylique secondaire (ou pentanol-2).

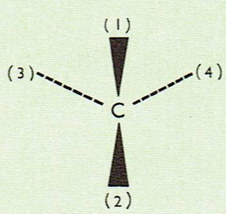
Pour obtenir les poids atomiques (ou atomes-grammes) portés sur ce tableau, on a posé d'abord, arbitrairement égal à 12 grammes exactement, celui de l'isotope le plus répandu du carbone (nombre de masse : 12). Le poids atomique correspondant à l'atome de carbone moyen des chimistes est alors 12,011 15, avec une erreur possible, en plus ou en moins, de 0,000 05. Si l'on veut connaître maintenant le poids atomique de l'oxygène, par exemple, il suffit de déterminer le poids de ce corps simple qui se combine à 12,011 15 g de carbone, pour former l'anhydride carbonique, de formule CO_2. Ce poids divisé par 2 donne la valeur cherchée : 15,999 4 ± 0,000 1. Avant 1961, on le posait arbitrairement égal à 16 g. La nouvelle convention entraîne donc de légères modifications dans les valeurs des poids atomiques. Les valences en traits pleins fins sont supposées dans le plan de la feuille; en pointillé, dirigées vers l'arrière; en triangle allongé (trait gras), dirigées vers l'avant.



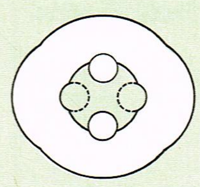
Molécule d'eau réelle : 1 cm représente 1 A. Les cercles en trait plein figurent les sphères de covalence. Les cercles pointillés figurent les sphères de Van de Waals qui déterminent l'encombrement de la molécule dans l'espace.



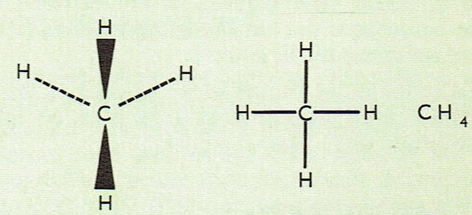
Représentations symboliques de la molécule d'eau.



Disposition dans l'espace des 4 valences qui partent d'un atome de carbone.



Molécule de méthane réelle.



Représentations symboliques de la molécule de méthane.

Les liaisons qui se trouvent en avant du plan figuré dans le tableau ci-dessus sont représentées par le signe en forme de triangle allongé ayant sa pointe dirigée vers l'arrière, tandis que les liaisons situées dans le plan du tableau sont figurées par des traits pleins et celles qui sont situées en arrière par des traits en pointillé.

Ces molécules, diverses par la nature de leurs atomes et aussi de plus en plus longues, sont loin d'être rigides; à l'instar des plus simples, des vibrations les animent. De plus, en solution, elles se déforment continuellement par rotation autour des liaisons simples, les angles qu'elles font entre elles restant constants. L'application de cette règle au propane ou au butane montre l'aspect changeant de leurs molécules; celles des polyéthylènes se comportent à peu près comme un fil de 2 à 3 Å de diamètre et de 10 000 à 100 000 Å de long, se recroquevillant en une pelote embrouillée constamment remaniée; grossi à notre échelle, ce fil aurait de 3 à 50 m de long pour 1 mm de diamètre.

UN NOUVEAU TYPE DE LIAISON : LES DOUBLES LIAISONS

Jusqu'à présent, nous avons lié l'une des valences d'un atome à une valence d'un autre, mais, fréquemment, elles se réunissent par paires, formant des doubles liaisons, comme dans une molécule très simple, l'éthylène.

La double liaison peut s'intégrer dans une chaîne d'une longueur quelconque, et plusieurs doubles liaisons figurent souvent sur une même chaîne; elle peut intéresser d'autres atomes que le carbone, tels l'oxygène ou l'azote. Lorsque la complication s'accroît, il devient nécessaire de repérer les différents atomes de carbone d'une chaîne au moyen d'un numéro d'ordre, en commençant par une extrémité.

L'ARCHITECTURE DES MOLÉCULES A DOUBLES LIAISONS

Les sphères de covalence de deux atomes réunis par une double liaison ne sont plus tangentes, mais empiètent légèrement l'une sur l'autre, si bien que la distance entre les centres de deux atomes de carbone, qui est dans l'éthane 1,54 Å, devient 1,33 Å dans l'éthylène.

Deux autres règles remplacent celles qui ont été énoncées plus haut : les valences partant des atomes doublement liés se trouvent dans un même plan, et toute rotation autour de la double liaison est impossible.

Ce sont des particularités de cette nature qui distinguent les deux butènes-2, et, nous le verrons, le caoutchouc de la gutta-percha.

DES MOLÉCULES BOUCLÉES EN CYCLE : LE BENZÈNE DE KEKULE

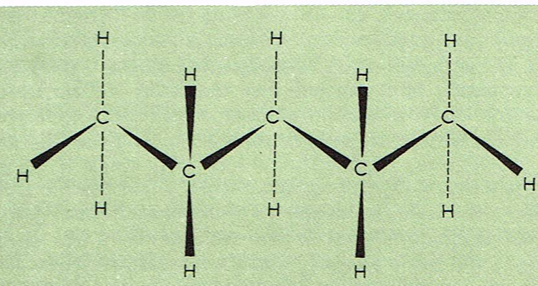
On conçoit, en appliquant les principes de la valence, qu'une chaîne d'atomes puisse se fermer sur elle-même, comme un collier; la molécule est dite alors « cyclique ». Les cycles les plus fréquents et les plus stables comptent 5 ou 6 atomes, mais beaucoup en contiennent 3, 4, 7, 8 ou un plus grand nombre.

Les centres des atomes de carbone du benzène, cycle à 6 atomes, se trouvent dans un même plan; **la formule de Kekule, reproduite ici, représente donc exactement l'aspect de cette molécule,** à une correction près; il est impossible, dans le cycle, de distinguer entre doubles liaisons et liaisons simples. La mécanique quantique retrouve ici un fait connu depuis longtemps des chimistes; c'est la raison pour laquelle ils ont convenu de représenter le cycle benzénique par un hexagone régulier, à l'intérieur duquel figure un cercle ou la lettre φ.

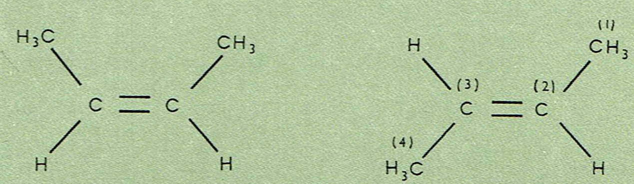
Nous avons vu les atomes se lier en chaînes, les chaînes se fermer en collier; elles peuvent aussi se ramifier, comme on le verra sur les formules de la page ci-contre.

une particularité de la chimie organique : la possibilité de molécules variées et complexes

Les particularités de structure qui viennent d'être exposées, et d'autres encore que nous verrons (atome de carbone asymétrique), peuvent évidemment figurer toutes ensemble dans une même molécule. **On entrevoit donc l'immensité du domaine que recouvre la chimie organique; tout l'imaginable ne peut être effectivement réalisé, mais le nombre des combinaisons isolées dépasse le million, et, chaque jour, les laboratoires répartis dans le monde entier en préparent de nouvelles.**

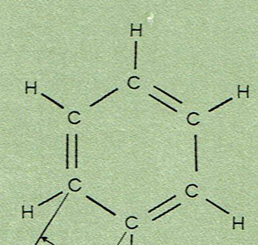


Molécule de pentane dans sa plus grande extension. Dans la réalité, elle se déforme constamment, car la rotation est libre autour des liaisons C—C. Représentation symbolique : CH_3—CH_2—CH_2—CH_2—CH_3.

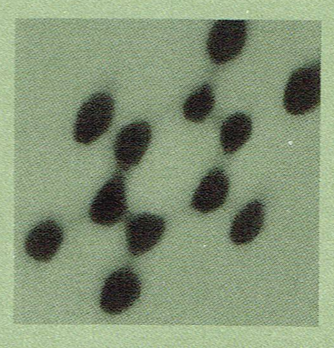


butène 2 cis — butène 2 trans

La double liaison fait partie ici d'une chaîne de quatre carbones. Toute rotation étant impossible autour d'elle, ces molécules correspondent à deux substances différentes.

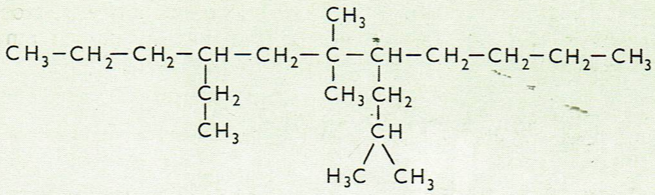


Formule du benzène de Kekule. Il ne savait pas que la molécule était plane et les six côtés de l'hexagone égaux.

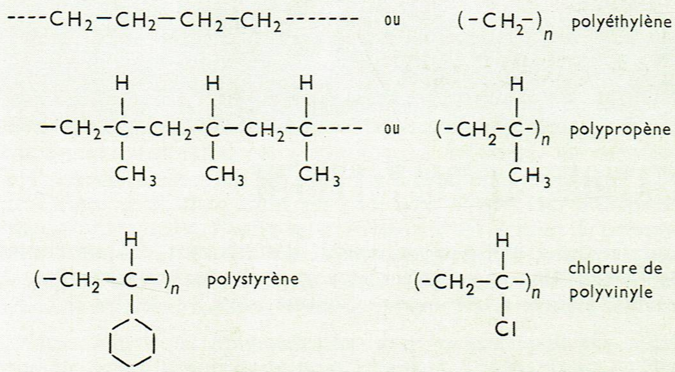


Hexaméthylbenzène vu aux rayons X et grossi 175 millions de fois. On distingue les 12 atomes de carbone (les 18 atomes d'hydrogène de la molécule sont invisibles).

LES HAUTS POLYMÈRES : DES MACROMOLÉCULES ENCORE SIMPLES



Exemple de molécule ramifiée.



(Quelques homopolymères usuels. Le motif élémentaire est répété *n* fois dans une macromolécule; *n* est grand et variable de l'une à l'autre.)

A B B A A A B A B A B B A B A B
A B B A A A B A B A B B A B A B
A B B A A A B A B A B B A B A B

Copolymère naturel.
Toutes les macromolécules sont rigoureusement identiques.

A B B A A A B A B A B B A B A A
B A B A A B A A B A A A B B A B B B A
A B B B A A B B B B A A A B B A A B A A B B A

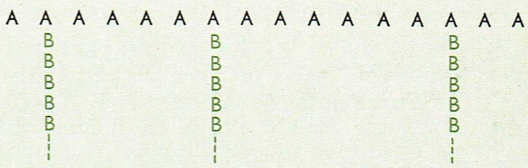
Copolymère statistique synthétique.
Les macromolécules diffèrent par l'ordre des motifs et par leur nombre.

A B A B A B A B A B A B

Copolymère alterné.

A A A A A A A A A A B B B B B B B B B B B B

Copolymère séquencé.



Copolymère greffé.
Des chaînes polymériques de motifs B partent de place en place, de la chaîne de motifs A.

Nous avons déjà utilisé le terme de « macromolécule », dont la signification n'offre plus de difficulté. La plupart de celles qui proviennent de synthèses, ou sont fournies par la nature, offrent une particularité de structure qui en simplifie notablement l'étude; elles sont formées par la répétition d'un, de deux ou de plusieurs groupements d'atomes, appelés « motifs élémentaires »; c'est ce qui justifie le nom de « hauts polymères » qui leur est donné.

Les molécules de polyéthylène résultent de la répétition d'un motif élémentaire unique et particulièrement simple, le méthylène — CH_2—. En fait, il serait plus correct de le nommer « polyméthylène ».

Les macromolécules construites à l'aide d'un seul motif élémentaire sont nommées « homopolymères ». Si deux ou plusieurs motifs élémentaires de natures différentes interviennent, on a des « copolymères ». Ils s'enchaînent suivant un ordre rigoureusement déterminé dans les macromolécules de certains produits naturels : protéines et acides nucléiques. L'art du chimiste ne peut prétendre à une telle perfection dans le détail des structures, si bien que les copolymères préparés au laboratoire sont des ensembles de macromolécules différant non seulement par le nombre total de leurs motifs et leurs proportions, mais aussi par leur ordre. S'il était possible de les voir et de les comparer, on aurait l'impression de se trouver devant de longues séries de numéros tirés à la roulette, d'où le nom de « copolymères statistiques » donné à ces substances.

Nous calculerons ultérieurement, à propos des acides nucléiques, le nombre de macromolécules distinctes, de longueur donnée, que l'on peut construire au moyen de deux ou plusieurs motifs, et montrerons que ce nombre fabuleusement élevé n'est pas sans intérêt pour la biologie.

Il est cependant possible d'obtenir par synthèse des copolymères présentant deux types d'ordre extrême. Le premier consiste en l'alternance de deux motifs, d'où le nom de « copolymères alternés »; il est clair qu'un tel produit, s'il était exempt d'irrégularités, se confondrait avec l'homopolymère de motif AB.

Dans le second type, celui des copolymères séquencés (ou à longues séquences), les motifs de chaque espèce forment des séquences plus ou moins longues et plus ou moins nombreuses.

L'étude des propriétés physico-chimiques de ces derniers ne fait guère que commencer.

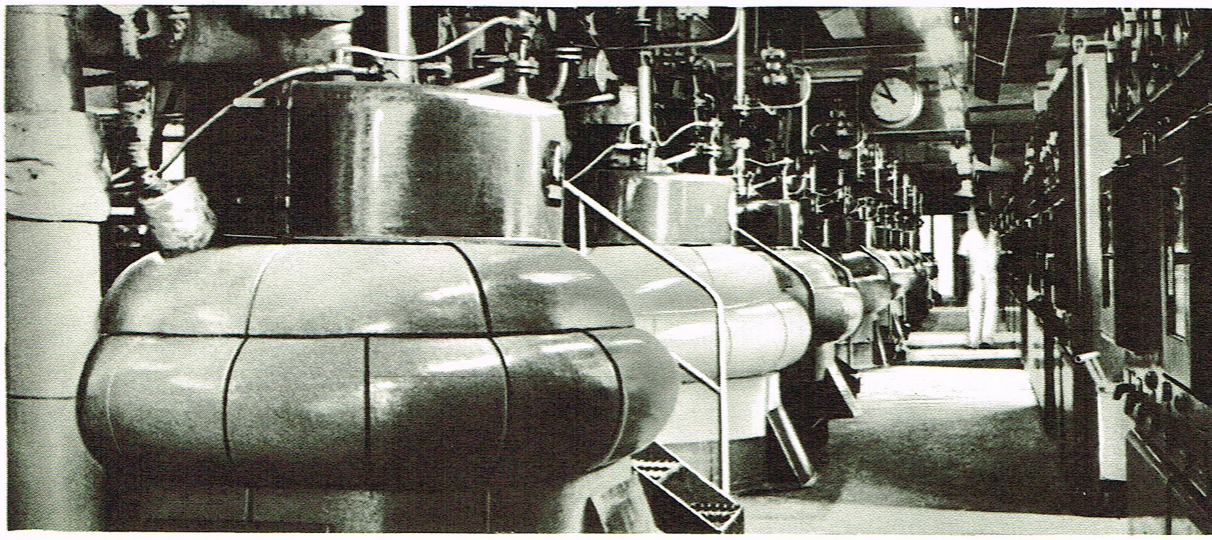
Assez voisins des copolymères séquencés, les copolymères greffés connaissent un grand succès; leurs macromolécules sont formées par une longue chaîne sur laquelle, par des moyens chimiques, ont été fixées des chaînes de même nature ou, la plupart du temps, de nature différente.

Nous n'avons pas épuisé toutes les particularités de structure des hauts polymères; de nouvelles se présenteront par la suite.

comment peuvent se terminer les chaînes macromoléculaires

En voyant les formules écrites plus haut, on remarquerait à juste titre qu'une valence libre figure à chaque extrémité. S'il en était bien ainsi, la stabilité des produits correspondants laisserait à désirer. Ces valences sont saturées par des atomes ou des groupements d'atomes dont la nature varie d'un mode de préparation à un autre. On serait tenté d'affirmer que ces groupements, de poids infime par rapport à tout le reste, sont en quelque sorte des impuretés des macromolécules, influençant de façon négligeable leurs propriétés. Cela n'est vrai que très approximativement; par exemple, il suffit que l'un des groupes terminaux soit acide pour que les colorants basiques se fixent sur lui.

Industriellement, cette particularité prend une certaine importance, car elle permet d'élargir les possibilités de coloration de certaines fibres synthétiques. De même, suivant la nature des extrémités de ses macromolécules, une substance peut résister plus ou moins vaillamment aux agents banals de dégradation : la chaleur, l'oxygène de l'air, etc.



Batterie d'autoclaves pour la fabrication du Nylon.

UN BEL EXEMPLE DE POLYMÉRISATION : LE POLYSTYRÈNE

Le polystyrène ne figure pas en tête des matières plastiques par le tonnage de sa production, mais c'est l'une des plus anciennes. Ses propriétés et les circonstances de sa préparation, à partir de son monomère, le styrène, ont fait l'objet des études les plus approfondies. Il est donc tout indiqué d'exposer, à propos de cet exemple, des notions fondamentales dont la valeur générale apparaîtra avec évidence.

du styrène aux polystyrènes

Le styrène (ou styrolène) est un liquide incolore, peu volatil, puisqu'il bout à 145 °C. Il suffit, cependant, que le flacon qui en contient reste quelques instants débouché, pour que sa présence se révèle par une odeur inquiétante de gaz d'éclairage. Celui-ci en recèle d'ailleurs des traces; combien de fois, dans les laboratoires, a-t-on recherché avec obstination une fuite de gaz inexistante, abusé par une goutte de styrène répandue par mégarde!

Ce produit, lorsqu'il est pur, perd lentement sa fluidité, pour devenir, au bout de quelques semaines, une masse visqueuse, puis un solide : le polystyrène. Comme la plupart des réactions chimiques, celle-ci s'accélère si l'on chauffe : à une température de 60 à 100 °C, par exemple, la même évolution se termine en quelques jours, voire en quelques heures.

Le bilan du phénomène s'exprime avec une grande simplicité; le styrène a fait place au polystyrène, sans que d'autres substances soient intervenues ou aient été produites simultanément. S'il nous était possible de suivre individuellement les molécules, nous les aurions vues s'enchaîner les unes aux autres pour former de longs chapelets qui peuvent en comporter de 2 000 à 100 000, et même plus encore; c'est en cela que consiste la polymérisation.

Tout se passe comme si chaque molécule qui la subit ouvrait l'une des deux liaisons (*a*), de telle sorte que chacun des deux atomes de carbone aurait une main libre pour saisir une main voisine.

Si l'on examine ce phénomène avec attention, on s'aperçoit que l'enchaînement d'un « motif élémentaire » au suivant peut se faire « tête à queue » ou « tête à tête ».

En fait, la première éventualité est la plus fréquente, mais la deuxième n'est pas toujours à négliger. Même si les enchaînements tête à tête sont rares, ils rompent de façon irrégulière la belle ordonnance des molécules de polystyrène, si bien que, à longueurs égales, deux d'entre elles ne sont jamais identiques. La formule abrégée fait donc abstraction, d'une part des irrégularités qui viennent d'être mentionnées, d'autre part des extrémités; la liaison libre que porte chacune d'elles est en effet, dans la réalité, saturée d'une façon ou d'une autre.

comment agissent les promoteurs

Une élévation de température n'est pas le seul facteur susceptible d'accélérer la polymérisation du styrène. On arrive à un résultat semblable en l'exposant à toutes sortes de radiations : la lumière ultraviolette, les rayons γ produits par le cobalt radio-actif de synthèse, des faisceaux de corpuscules matériels de toute nature — électrons, neutrons, noyaux d'atomes. Ce sont là des moyens que l'on peut qualifier de « physiques »; non moins puissant est celui qui consiste à introduire dans le liquide une proportion infime de certaines substances bien définies. Les chimistes les classent en trois grandes catégories : les générateurs de radicaux libres, les bases et les acides. On les qualifie de « promoteurs », pour les distinguer des catalyseurs vrais; en effet, contrairement à ces derniers, ils sont détruits dans la mesure même où ils provoquent la réaction considérée.

les radicaux libres, particules actives et insaisissables

Au début de ce chapitre, nous avons mentionné à plusieurs reprises comment devaient être saturées les valences au sein d'une molécule; mais il s'agissait de molécules en général très stables, et que l'on pourrait dire normalement constituées. Déjà vers 1900, les travaux de Gomberg obligèrent à admettre l'existence du triphénylméthyle, qui, à l'époque, passait pour un monstre, puisque l'un de ses carbones porte une valence libre. On lui donne pour cela le nom de « radical libre ».

Pressés par les faits, et de plus en plus hardis, les chimistes ont admis que des radicaux libres très actifs, c'est-à-dire instables et fugitifs, se forment au cours de nombreuses réactions. Ce sont des particules en somme assez banales, mais insaisissables, dont l'existence se manifeste seulement par les effets qu'elles produisent.

Il en naît lorsqu'on chauffe du peroxyde de benzoyle; leur existence éphémère se termine par leur transformation en molécules stables. Si du styrène est présent, le sort d'une certaine proportion de ces radicaux change; ils attaquent les molécules de styrène et se fixent sur l'un des atomes de carbone doublement liés.

Phot. Blanc et Demilly.

Un nouveau radical en résulte, lui aussi chimiquement très actif, qui attaque une nouvelle molécule de styrène, et ainsi de suite. Un seul radical initiateur provoque l'enchaînement bout à bout d'un nombre considérable de molécules du monomère, cela jusqu'à ce que, au hasard des rencontres, un de ces longs radicaux perde sa valence libre, c'est-à-dire sa grande réactivité chimique. Cela arrive lorsque les extrémités actives de deux macroradicaux se rencontrent, et que les deux valences libres se soudent. D'autres processus, qui ne seront pas décrits ici, peuvent aboutir à la même conclusion.

En l'absence de peroxyde de benzoyle, et de tout autre promoteur, il est probable que la polymérisation du styrène s'effectue selon un mécanisme voisin. Dans une première étape se forme alors un radical actif, par choc de deux molécules de styrène. Toute réaction engendrant des radicaux libres agressifs entraînera la polymérisation du styrène, et ces réactions sont légion.

Tous les peroxydes se comportent à peu près comme le peroxyde de benzoyle, et certains sont, comme lui, utilisés tant au laboratoire qu'industriellement.

Lorsque la polymérisation doit s'effectuer au voisinage de la température ordinaire, on a avantage à mélanger, au moment de l'emploi, deux produits séparément stables, réagissant l'un sur l'autre avec formation de radicaux libres. Généralement, l'un est un oxydant, l'autre un réducteur, d'où le nom de promoteurs « redox ». Citons, parmi ces couples très nombreux, le mélange d'eau oxygénée et de sulfate ferreux, qui engendre des radicaux oxhydryles libres, — OH.

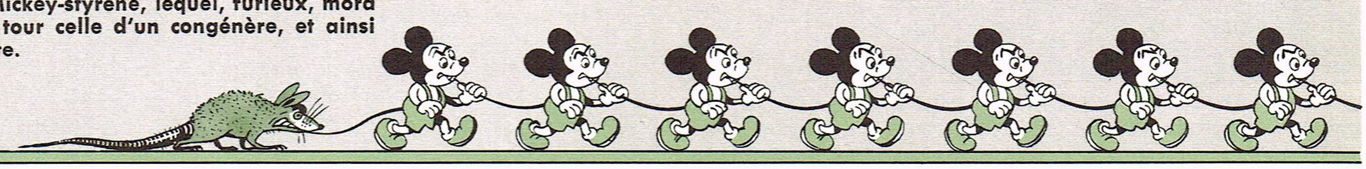
D'autres agents de polymérisation sont encore employés : des produits acides (fluorure de bore, chlorure d'aluminium) ou des produits basiques (métaux alcalins, composés organométalliques). Les catalyseurs de Ziegler-Natta conduisent à une variété isotactique du polystyrène, comme nous le verrons.

les agents qui modifient le cours d'une polymérisation

Le styrène se polymérise donc très facilement, ce qui ne laisse pas d'être inquiétant pour son bon état de conservation en cours de stockage. On y arrive en l'additionnant d'un stabilisateur, ou inhibiteur. Une telle substance réagit sur les radicaux beaucoup plus rapidement que le styrène, même si elle est ajoutée dans une faible proportion : elle les capte aussitôt formés, les empêchant de provoquer une polymérisation dans des circonstances où on la juge indésirable. La quinone, la phényl-β-naphtylamine, à la concentration de 1 pour 1 000, conviennent très bien à cet usage.

D'autres produits, en tête desquels viennent les dérivés nitrés, dinitrobenzène et autres, diminuent la vitesse de polymérisation sans l'annuler; eux aussi se combinent aux radicaux, mais en en donnant d'autres, plus « paresseux ».

De nombreux solvants, parmi lesquels le tétrachlorure de carbone, et, à faible dose, les mercaptans, agissent de façon différente; ils ne modifient guère la vitesse de polymérisation, mais les macromolécules de polystyrène formées sont plus courtes. Un macroradical, en effet, interrompt sa croissance en arrachant un atome à ces substances appelées « agents de transfert »; il en résulte un nouveau radical. Mais celui-ci est très actif, et non plus paresseux, comme dans le cas des retardateurs. Immédiatement, il provoque une nouvelle polymérisation.



Symbole de la polymérisation. Un rat (radical libre promoteur) mord la queue d'un Mickey-styrène, lequel, furieux, mord à son tour celle d'un congénère, et ainsi de suite.

LES PROCÉDÉS INDUSTRIELS DE PRÉPARATION DU POLYSTYRÈNE

Une considération essentielle détermine le choix d'un procédé industriel : le prix de revient. Il semblerait donc inutile de modifier la technique de préparation du polystyrène développée plus haut. Additionner le monomère de 0,5 p. 100 de peroxyde de benzoyle, chauffer à 70 °C et attendre quelques heures la fin de la réaction, imaginerait-on opérations plus simples? *S'en tenir là serait négliger certains petits détails, qui, lorsqu'on traite des quantités de matières premières de l'ordre de plusieurs tonnes, prennent une importance considérable.*

Les réactions chimiques s'effectuent en général avec un appréciable dégagement de chaleur, selon la vieille règle de Berthelot, qui, pour approximative qu'on la tienne, conserve toujours sa valeur. Or, l'intégration de 10 molécules-grammes de monomère, autrement dit la formation de 1 kg de polymère, dégage 160 grandes calories, ce qui suffirait à élever de 160 °C la température de 1 litre d'eau. Un dispositif permettant de réfrigérer la masse s'impose donc.

Il est un inconvénient non moins grave : à la fin, on se trouve en présence d'un solide de plusieurs mètres cubes, qu'il faut extraire de l'autoclave.

la polymérisation en perles

On y parvient, mais un moyen simple et ingénieux permet de tourner la dernière difficulté, tout en facilitant la régulation de la température. Il consiste à briser, diviser la masse du styrène, avant qu'elle ait évolué vers l'état solide, en l'agitant vigoureusement avec une quantité d'eau allant de un à quatre fois son propre volume. Le peroxyde de benzoyle, promoteur peu soluble dans l'eau, reste dans le monomère et accomplit son œuvre. L'agitation divise le styrène en gouttelettes liquides, qui, progressivement, deviennent des grains solides glissant les uns sur les autres, si toutefois, au cours de l'évolution du système, ils ont échappé à un danger qui les guette : l'agglomération en paquets compacts. On l'évite en ajoutant un stabilisateur (sulfate de baryum, carbonate ou phosphate de calcium, amidon, albumine, etc.).

Ce procédé est qualifié de « polymérisation en perles ». A la fin, les perles sont extraites de l'autoclave, lavées, séchées et empaquetées.

On conçoit que si la même chaleur dégagée se répand dans un volume de liquide beaucoup plus grand et de chaleur spécifique plus élevée, la température monte moins rapidement; d'autre part, si le milieu restant est agité vigoureusement jusqu'à la fin, l'efficacité du système réfrigérateur est plus grande.

la polymérisation en émulsion

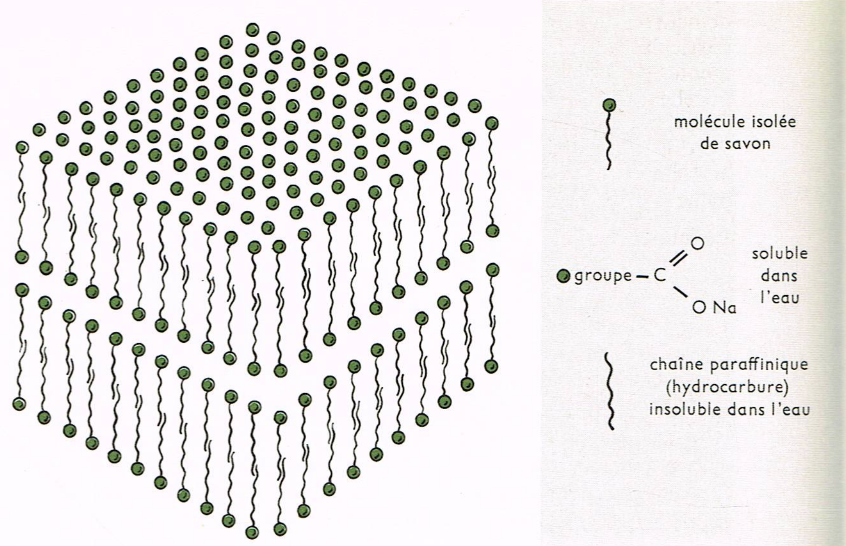
Des avantages analogues s'attachent à la polymérisation en émulsion. Celui qui assisterait pour la première fois à l'opération aurait tendance à parier qu'elle fournirait des résultats identiques à la polymérisation en perles. Elle débute de la même façon, à deux petits détails près : le stabilisateur ajouté au mélange continuellement agité de styrène et d'eau n'est plus du phosphate de calcium, mais du savon de Marseille; le promoteur choisi, le persulfate de potassium, se dissout dans l'eau (contrairement au peroxyde de benzoyle) et ne pénètre pas dans le styrène. Mais, en fin de polymérisation, à l'ouverture de l'autoclave, le pari est perdu. Au lieu de petites billes solides baignant dans l'eau, on distingue un latex opalescent, rappelant par son aspect le liquide, chargé en globules de caoutchouc, qui coule des entailles pratiquées dans l'écorce d'hévéa. Un examen au moyen du puissant microscope électronique y révèle des sphères minuscules de polystyrène, dont le diamètre varie de 500 à 8 000 Å.

De nombreux chercheurs ont tenté de déceler le mécanisme de la formation de ce latex. Son trait essentiel est que la polymérisation ne commence pas dans de grosses gouttes de monomère, qui se solidifieraient en conservant leur taille géante, mais dans d'infimes amas de molécules de styrène; eux-mêmes ont été captés par des amas guère plus volumineux de molécules de savon.

Il ne faut pas croire, en effet, que les molécules de savon restent isolées les unes des autres lorsqu'on les dissout dans l'eau. Elles s'associent en groupes de 50 à 100 unités, appelés « micelles ». La forme exacte de ces micelles est mal connue, mais les traits fondamentaux de leur structure ne font pas de doute. Ils sont fort compréhensibles si l'on examine une molécule de savon isolée, par exemple celle du stéarate de sodium. Elle comprend une chaîne de 17 atomes de carbone, analogue à celle d'une paraffine, que termine à une de ses extrémités un groupement carboxylate de sodium.

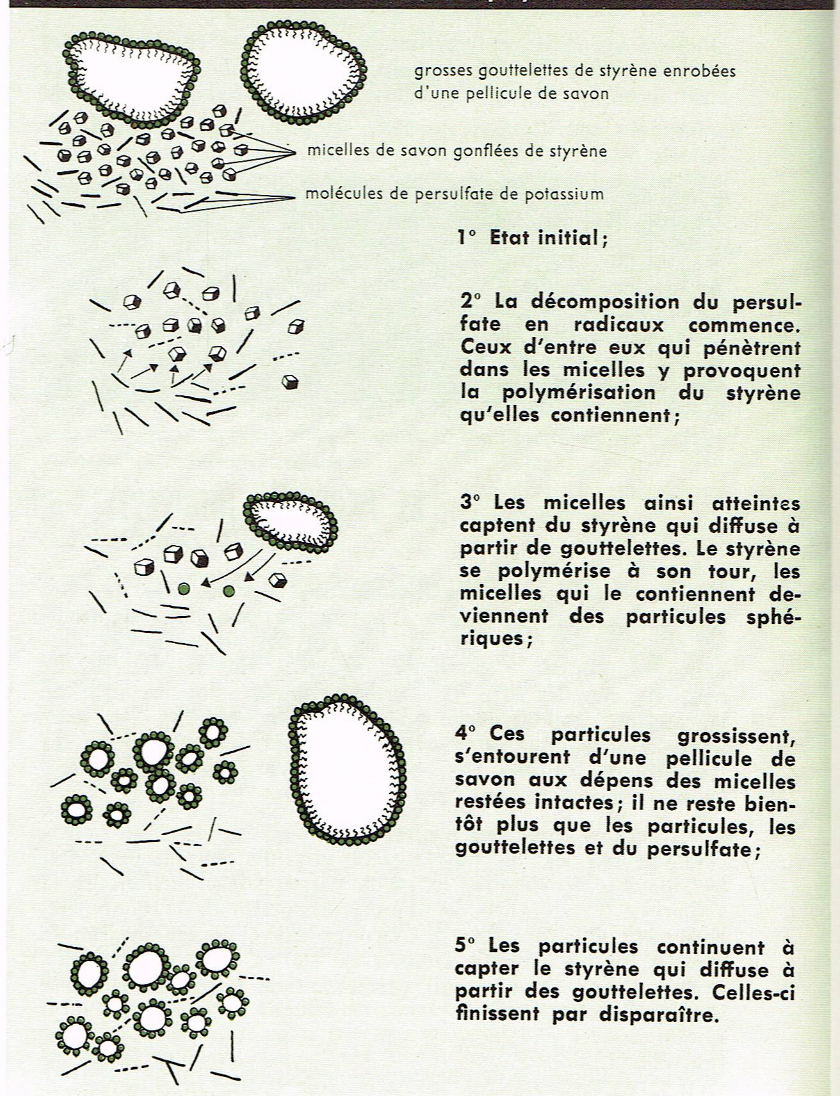
La chaîne des 17 carbones, insoluble dans l'eau comme la paraffine, se trouve entraînée dans ce liquide par le groupement carboxylate soluble, mais avec regret. Aussi recherche-t-elle la compagnie de ses congénères. Il en résulte la disposition de la figure, c'est-à-dire de petits disques dont les faces hydrophiles baignent dans l'eau, tandis que l'intérieur possède les propriétés d'un minuscule morceau de paraffine, ou d'une huile minérale. A ce titre, le styrène va s'y dissoudre jusqu'à saturation, au début de la polymérisation en émulsion, lorsqu'on agite ce monomère dans la solution de savon. Un centimètre cube du liquide aqueux contient alors par exemple 10^{18} micelles. Chacune est formée de 100 molécules de savon emprisonnant 100 molécules de monomère. L'excès de celui-ci est dispersé en gouttes géantes, qui joueront ultérieurement le rôle de réservoir. Tournons-nous maintenant vers le persulfate de potassium promoteur. Dissous dans l'eau, il se décompose lentement en engendrant des radicaux, notamment l'oxhydryle, — OH, dont la durée de vie est suffisante pour que chacun, pénétrant dans une micelle, y provoque le cataclysme que nous connaissons bien, c'est-à-dire la polymérisation du styrène qu'elle contient. Alors l'équilibre se trouve rompu, le polystyrène s'imbibe de nouvelles quantités de son monomère, puisé dans les gouttes géantes; la micelle grossit et devient une particule sphérique dont le volume continue à croître, grâce à l'apport incessant de monomère et de radicaux promoteurs.

Aurons-nous donc finalement autant de particules dans le latex qu'il existait, au début, de micelles de savon? Loin de là, car un phénomène dont nous n'avons pas encore parlé réduit progressi-



Coupe d'une micelle formée par la réunion de molécules de savon, au sein d'une solution aqueuse. Les micelles peuvent capter du styrène dans leur partie paraffinique.

Formation du latex au cours de la polymérisation en émulsion.



1° Etat initial;

2° La décomposition du persulfate en radicaux commence. Ceux d'entre eux qui pénètrent dans les micelles y provoquent la polymérisation du styrène qu'elles contiennent;

3° Les micelles ainsi atteintes captent du styrène qui diffuse à partir de gouttelettes. Le styrène se polymérise à son tour, les micelles qui le contiennent deviennent des particules sphériques;

4° Ces particules grossissent, s'entourent d'une pellicule de savon aux dépens des micelles restées intactes; il ne reste bientôt plus que les particules, les gouttelettes et du persulfate;

5° Les particules continuent à capter le styrène qui diffuse à partir des gouttelettes. Celles-ci finissent par disparaître.

vement leur nombre; les molécules de savon quittent les micelles demeurées intactes pour se fixer avec énergie sur la surface des particules en cours de croissance, et, comme cette surface augmente considérablement, il arrive un moment où toutes les micelles ont disparu. Les particules continuent à croître en absorbant du styrène, qui se polymérise, mais leur nombre, de l'ordre de 10^{14} par centimètre cube, reste alors constant. Lorsque la dernière gouttelette du styrène en réserve est épuisée, le phénomène s'arrête.

Le latex est enfin additionné d'un produit tel que l'hydroquinone ou le diméthyl-dithiocarbamate de potassium, qui bloque les radicaux libres et détruit le promoteur, précaution nécessaire en raison des transformations parasites qu'ils pourraient engendrer.

Ensuite, on procède à la coagulation. L'addition d'eau salée suffit à la provoquer; elle est complétée par de l'acide sulfurique dilué. Il faut ensuite filtrer, laver et sécher le produit.

Les différents procédés de polymérisation du polystyrène ont été adaptés à une fabrication continue.

PURIFICATION ET FRACTIONNEMENT DU POLYSTYRÈNE

Ceux qui ont pratiqué si peu que ce soit la chimie organique savent qu'un produit brut de préparation n'est jamais pur; pour le débarrasser des substances étrangères, il est nécessaire de le soumettre à des opérations plus ou moins compliquées, parmi lesquelles la cristallisation et la distillation sont les plus communes. Les composés macromoléculaires contiennent eux aussi des impuretés diverses, ne seraient-ce que les réactifs ajoutés pour provoquer la polymérisation, les savons, agents de transfert, etc. De tous ceux-là, on se débarrasse assez facilement, mais il est un fait beaucoup plus grave : les macromolécules d'un échantillon de polystyrène ne peuvent avoir la même longueur. C'est là une remarque de bon sens, confirmée par l'expérience.

Le problème se pose de trier les macromolécules suivant leur longueur, en d'autres termes de fractionner le produit brut. En effet, la plupart des études scientifiques nécessitent des échantillons aussi « purs » que possible, c'est-à-dire de répartition serrée autour d'une moyenne.

La technique employée repose sur cette loi simple que la solubilité d'un haut polymère est d'autant plus petite que sa masse moléculaire est plus grande. Nous n'entrerons pas dans le détail des opérations, qui sont longues, assez onéreuses, et nécessitent des précautions rigoureuses; en principe, le styrène est dissous dans le benzène et précipité progressivement par addition de quantités croissantes d'alcool méthylique. Les différentes fractions obtenues diffèrent d'emblée par leur aspect.

Les premières, de masses moléculaires élevées, se présentent en filaments tenaces, tandis que les dernières sont pulvérulentes.

La qualité d'un stock de fabrication industrielle dépend non seulement de sa masse moléculaire moyenne, mais aussi de sa répartition lâche ou serrée autour de celle-ci; le fractionnement peut servir de procédé d'analyse et de contrôle.

LES MULTIPLES USAGES DU POLYSTYRÈNE

Le polystyrène est livré aux industries de transformation en poudre ou en granules. Il fond vers 180 °C sans décomposition rapide, ce qui permet d'en fabriquer divers objets par injection du liquide dans des moules; ses usages sont multiples, allant des boîtes et bocaux aux revêtements muraux. Comme il est transparent et incolore, des colorants peuvent lui être incorporés. Une forme plus nouvelle, et dès maintenant populaire, de ce polymère est le polystyrène « mousse », ou encore « expansé », ce dernier terme n'étant que la transposition en français d'un mot anglais.

Par exemple, au lieu de styrène pur, on soumet au procédé de polymérisation en suspension un mélange de styrène et de 10 à 20 p. 100 d'éther de pétrole. Les perles obtenues retiennent ce dernier; introduites dans un moule chauffé à la vapeur, elles se gonflent et deviennent autant de gros grains semblables au « pop-corn », qui s'agglomèrent entre eux. On les distingue d'ailleurs parfaitement sur les objets fabriqués de cette façon.

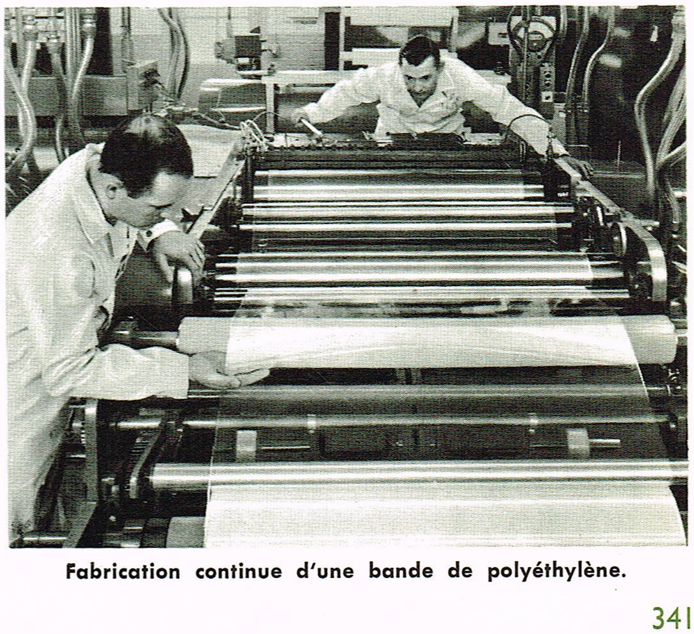
Le polystyrène mousse est blanc, brillant; en flocons, il imite la neige; quand il est répandu sur les sapins de Noël, l'illusion est parfaite. Par moulage ou extrusion, on lui donne les formes les plus diverses, par exemple celle de planches rigides. A volume égal, il pèse environ 30 fois moins que l'eau.

De tous les isolants thermiques, l'air est le meilleur lorsqu'il n'est pas agité; c'est pourquoi les mousses de styrène conduisent très mal la chaleur; elles sont également rigides, imperméables à l'eau liquide ou à la vapeur. Leur usage est tout indiqué pour calorifuger les plafonds, les murs et les tuyaux; elles sont d'ailleurs vendues débitées en planches ou en cylindres, que l'on travaille facilement et transporte sans effort. Des objets fragiles, œufs, appareils en verre, défient la casse lorsqu'ils sont encastrés dans des blocs de polystyrène mousse, creusés de façon adéquate.

Rappelons que l'un des succédanés les plus répandus du caoutchouc naturel, le G. R. S., est un copolymère butadiène-styrène. Le produit brut, non coagulé, de sa préparation en émulsion est commercialisé sous la dénomination « latex » et trouve un ample usage en peinture.

Enfin, ainsi que nous le verrons, le polystyrène, formé extemporanément en mélange avec des résines maléiques, réticule ces dernières en leur donnant une solidité comparable à celle du bois ou des métaux.

Offrant une bonne résistance à un grand nombre d'agents chimiques acides et basiques, le polystyrène se gonfle et se dissout



Fabrication continue d'une bande de polyéthylène.

Phot. Eastman chemical prod.

cependant dans divers liquides tels que le benzène; il faut prendre garde de le chauffer au-dessus de 150 à 180 °C, car il fond; le styrène mousse, en particulier, s'aplatit irrémédiablement. Malheureusement, enfin, il n'a guère été retenu comme textile artificiel, malgré l'aspect filamenteux qu'il prend lorsque sa masse moléculaire est élevée.

Ce que nous avons dit des techniques de polymérisation du styrène, du fractionnement du polymère et des formes de sa commercialisation pourrait s'appliquer, à des variantes près, à la polymérisation de beaucoup de monomères (chlorure de vinyle, méthacrylate de méthyle, nitrile acrylique) et aux produits qui en résultent.

DE NOUVELLES RÉGULARITÉS DANS LES CHAINES POLYMÈRES HONORÉES PAR LE PRIX NOBEL 1963

Si, au lieu de provoquer la polymérisation du styrène par des radicaux libres ou par les promoteurs courants, acides ou basiques, connus depuis longtemps, on utilise à cet effet le produit solide mal défini et certainement complexe obtenu en mélangeant du triéthyl-aluminium au trichlorure de titane, le polystyrène formé possède une propriété nouvelle : il cristallise. Cette cristallisation, observable par la diffraction des rayons X, se traduit par une cohésion plus grande entre les macromolécules, et entraîne, à ce titre, d'autres caractères non moins manifestes. Une telle variété de polystyrène reste en effet insoluble dans l'heptane ou le benzène, ce qui d'ailleurs permet de la séparer; elle résiste aussi beaucoup mieux à la traction.

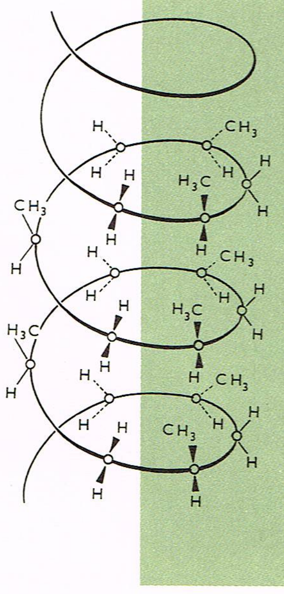
Quelle différence existe-t-il, aux yeux des chimistes, entre le polystyrène résolument amorphe et le produit de même nom cristallisable? Elle est suffisamment subtile pour demander à être examinée avec attention. Pour observer les détails de structure que nous avons en vue, donnons à la longue chaîne des atomes de carbone la forme d'une hélice, ou plutôt d'une ligne brisée inscrite dans une hélice. En effectuant cette opération, il va sans dire que nous modifions l'aspect général de la macromolécule par des rotations autour des liaisons simples, sans jamais exercer d'effort intempestif qui déformerait les angles que les valences issues d'un même atome font entre elles.

De chaque atome de carbone partent deux valences inscrites dans l'hélice; deux autres lient l'atome considéré soit à deux atomes d'hydrogène, soit à un atome d'hydrogène et à un noyau benzénique. Situées dans un plan passant à peu près par l'axe de l'hélice, elles se dirigent vers l'extérieur de celle-ci.

Considérons les carbones liés à la fois à un H et à un φ. Pour chacun, deux dispositions sont possibles : φ en haut et H en bas, ou bien l'inverse. Si tous présentent la même disposition, par exemple avec tous les φ en haut, on dit que le polymère est « isotactique ». C'est une structure très particulière; en général, les deux dispositions se succèdent dans un grand désordre; le polymère est alors qualifié d' « atactique ». On conçoit que la structure des chaînes isotactiques permette un arrangement régulier de celles-ci les unes par rapport aux autres, ce qui est synonyme de « cristallinité ».

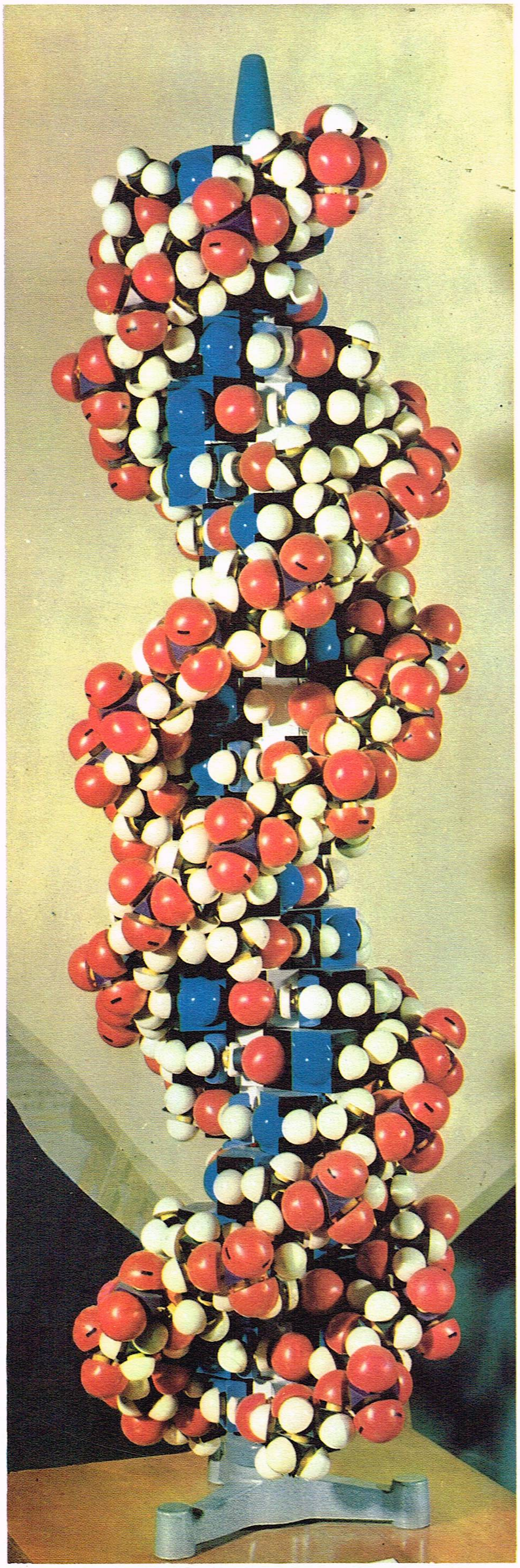
D'autres dispositions régulières de φ et H son concevables, parmi lesquelles le type syndiotactique a pu être réalisé synthétiquement; là, alternativement, φ se trouve en haut, puis en bas. Ces considérations ne sont évidemment pas liées à la nature du noyau benzénique, φ; elles s'appliquent à tous les polymères de même type, comme le propène par exemple. Le polypropène isotactique, commercialisé sous le nom de « Meraklon », forme des fils dont la solidité est comparable à celle des meilleurs Nylons (charge à la rupture : environ 75 kg pour un fil de 1 mm² de section).

Les promoteurs susceptibles d'engendrer des isotactiques ou des syndiotactiques sont en nombre considérable, mais, pour la plupart, du même type que celui qui a été indiqué plus haut. Ils possèdent également la propriété de provoquer, à température et à pression peu élevées, la polymérisation de l'éthylène, en donnant un polyéthylène dont les macromolécules ne portent pas de ramifications. Pour cette raison, c'est également un produit cristallisé partiellement, dont les qualités mécaniques surclassent celles des polyéthylènes fabriqués par des méthodes plus anciennes. Les mêmes promoteurs sont également capables de provoquer la transformation du butadiène ou de l'isoprène en toutes les variétés pures possibles de leurs polymères. Les noms de Ziegler et de Natta, qui ont reçu conjointement le prix Nobel de chimie en 1963, leur sont associés.



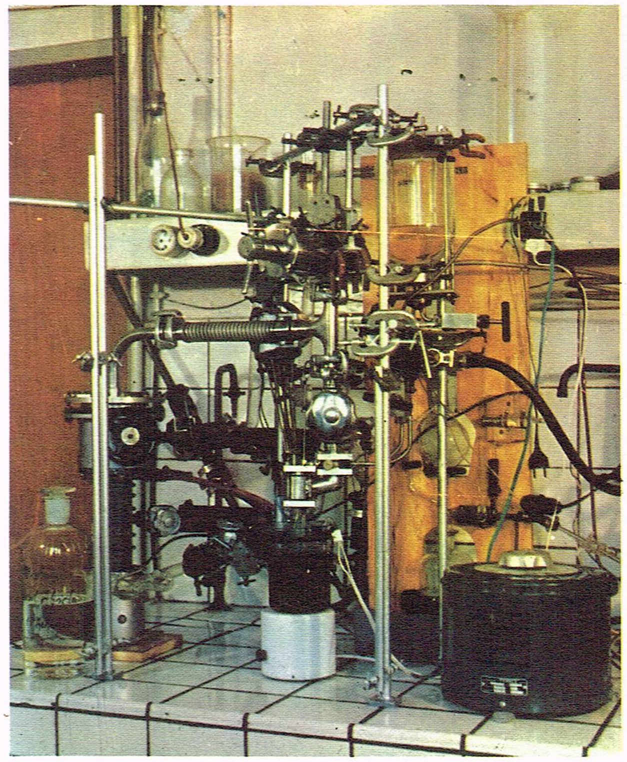
Polyisopropène isotactique (Meraklon).

Les cercles sur l'hélice représentent les atomes de carbone de la chaîne du polymère. Tous les CH_3 sont placés au-dessus de la spire à laquelle ils sont liés.

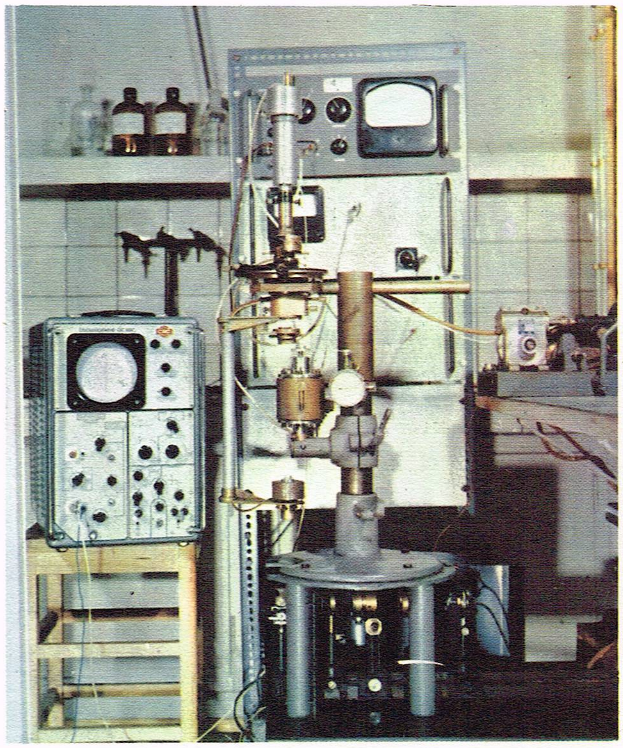


Aux sources mêmes de la vie : l'A. D. N., dont la molécule ne comporte pas moins de 10 000 motifs élémentaires.

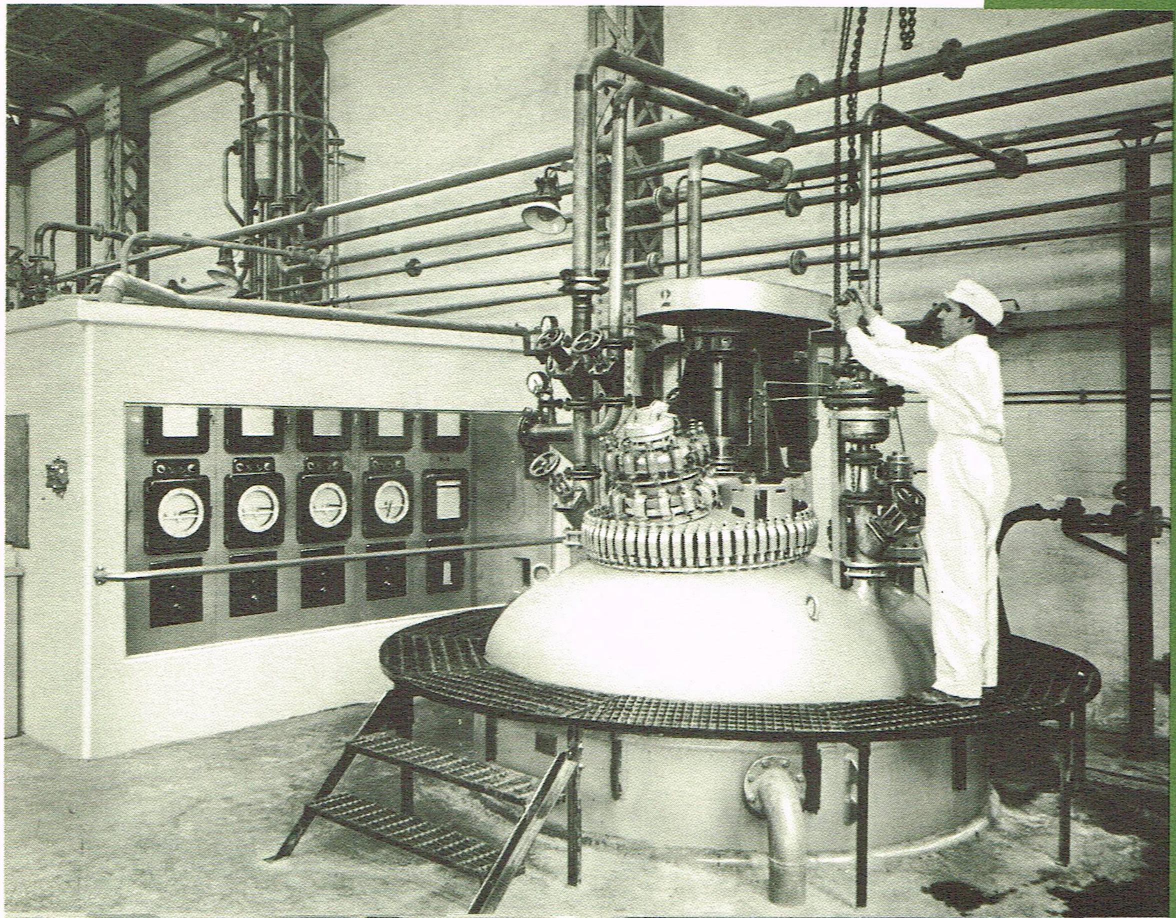
Phot. Larousse.



L'étude des macromolécules met en œuvre d'innombrables techniques. Les deux installations représentées ici servent à étudier la polymérisation des diènes *(ci-dessus)* et la biréfringence *(ci-dessous)*. *Phot. C. N. R. S.*

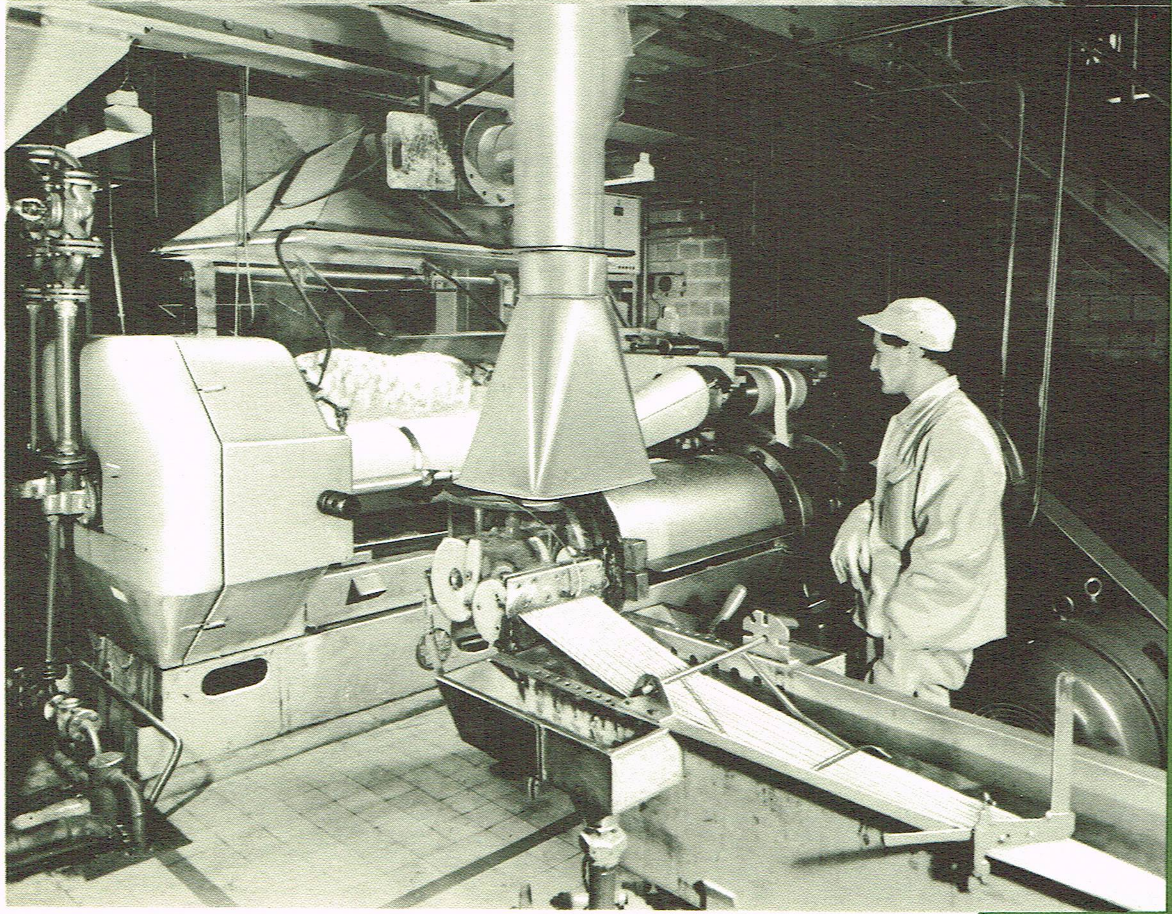


Phot. Larousse.



Fabrication du polystyrène.

Autoclave de polymérisation.



La masse de polystyrène est « calandrée » et poussée sous pression par les trous d'une filière.

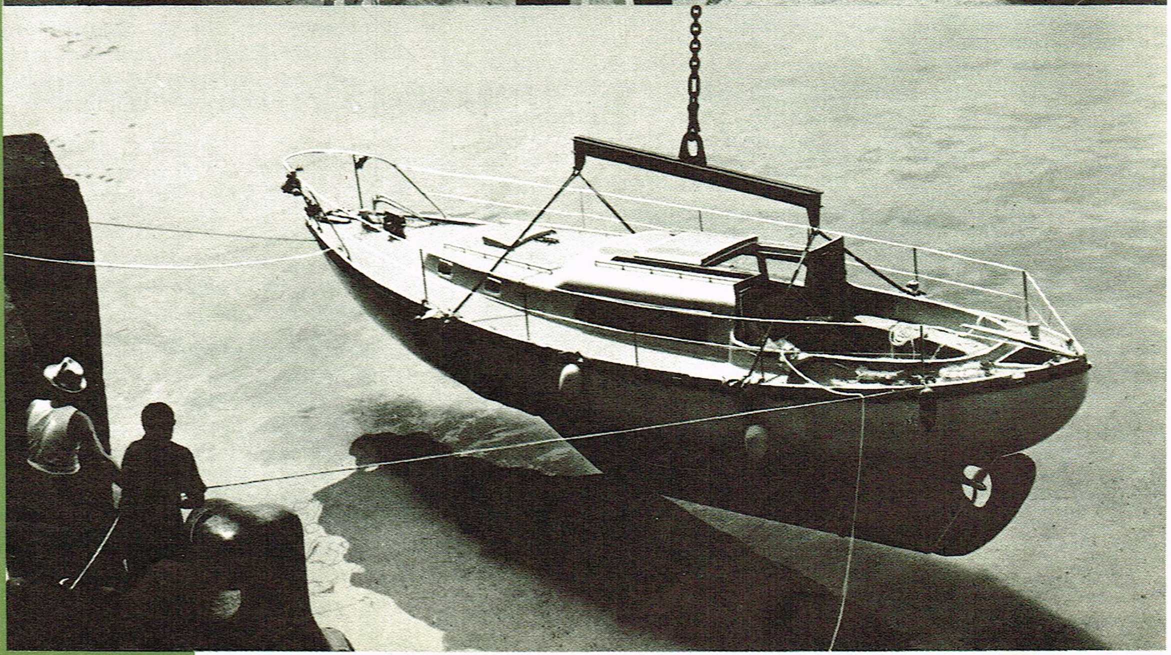
Doc. Ets Kuhlmann.

Photothèque.





Des pièces moulées les plus compliquées aux coques des bateaux faites en petites séries, il n'est pas de domaine où les matières plastiques ne trouvent des applications.



Doc. Saint-Gobain-Péchiney.

Doc. Lanaverre.

DU CAOUTCHOUC NATUREL AU CAOUTCHOUC DE SYNTHÈSE

Le caoutchouc, substance au comportement étonnant, intimement lié à la vie contemporaine, était connu des indigènes de l'Amérique centrale, qui savaient traiter le latex d'hévéa; ils en fabriquaient des balles et d'autres petits objets lorsque les Européens découvrirent ce continent. Mais, pour qu'on parvînt à le produire synthétiquement et à jeter quelques lueurs sur le processus biologique qui l'engendre, il fallut attendre ces toutes dernières années.

COMMENT RÉALISER LA SYNTHÈSE DU CAOUTCHOUC

Les macromolécules de caoutchouc naturel se construisent par la répétition d'un motif élémentaire dérivé de l'isoprène; ce dernier, que l'on prépare synthétiquement, se polymérise comme le styrène, sous l'effet des mêmes promoteurs. Mais on obtient alors, suivant la nature de ceux-ci, des produits dont la variété apparaît lorsqu'on se livre à une étude minutieuse de leurs propriétés; leur simple aspect les rangerait dans la catégorie assez vague des substances caoutchouteuses.

Pour comprendre la raison de cette diversité, et, du même coup, préciser la structure du caoutchouc naturel, il suffit de se livrer sur l'isoprène au jeu de construction que nous avons pratiqué sur le styrène; sa complication sera seulement un peu plus grande.

La molécule d'isoprène est reproduite ci-dessous, avec le numérotage habituel des atomes de carbone.

On aperçoit d'abord deux séries possibles de polymères, suivant que la double liaison qui s'ouvre est placée en 1-2 ou en 3-4.

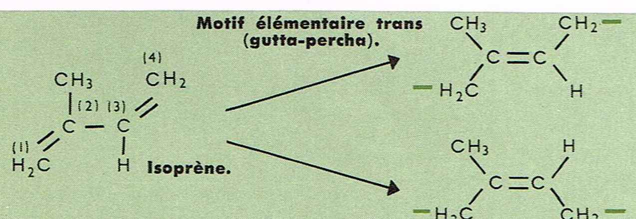
Mais les substances précédentes ne sont en aucune façon du caoutchouc naturel. C'est que, avec l'isoprène, apparaissent les modes de polymérisation 1-4 cis et 1-4 trans, qui ne trouvent pas d'équivalent pour le styrène. Ils sont dus au fait qu'une liaison simple, et une seule, sépare les doubles liaisons de l'isoprène.

entre caoutchouc et gutta-percha, une minime différence géométrique

Pour comprendre clairement ces nouveaux modes, ouvrons simultanément ces deux doubles liaisons.

Réunissons ensuite les liaisons libres médianes 2 et 3, ce qui donne l'une ou l'autre des structures « cis » ou « trans ». Leur répétition un grand nombre de fois correspond à deux sortes de produits, respectivement le caoutchouc et la gutta-percha.

Il aura peut-être échappé au lecteur quelles particularités distinguent « cis » de « trans ». Pour les comprendre, on se reportera à ce qui a été exposé précédemment : lorsque deux atomes de carbone sont réunis par une double liaison, une rotation autour de celle-ci est impossible; or, c'est l'opération qu'il faudrait effectuer pour passer d'un de ces motifs élémentaires à l'autre. Dans la disposition « trans », les carbones 1 et 4 se trouvent placés de part et d'autre de la double liaison 2-3, tandis qu'ils se trouvent du même côté par rapport à celle-ci dans la disposition « cis ».



Motif élémentaire cis (caoutchouc). [La répétition de ces motifs engendre des macromolécules très différentes, car toute rotation autour de la double liaison est impossible.]

La plupart des promoteurs de polymérisation, appliqués à l'isoprène, engendrent des successions désordonnées de tous les motifs élémentaires possibles. Les produits obtenus sont tout au plus des succédanés du caoutchouc.

La chance a résolument tourné lorsqu'on a expérimenté le lithium. C'est un métal ressemblant beaucoup au sodium et aux autres corps simples de la même famille (potassium, rubidium, césium), à la fois par ses propriétés et par celles de ses dérivés. Tous ces métaux, fraîchement coupés, présentent une surface très brillante, qui se ternit rapidement à l'air, car elle est oxydée par l'oxygène et attaquée par la vapeur d'eau. Le lithium se distingue cependant par quelques propriétés particulières : il est attaqué par l'azote et se recouvre presque instantanément, dans ce gaz, d'une couche noire d'azoture, tant et si bien qu'il faut le manipuler dans l'argon, le moins onéreux des « gaz rares ».

Des fragments de lithium ou de sodium plongés dans l'isoprène liquide s'entourent lentement de polymères. Jusqu'en 1956, on avait étudié surtout ceux qui se forment à la surface du sodium; nous imaginons la joie des chercheurs qui examinèrent pour la première fois attentivement le polymère recouvrant le lithium et découvrirent son identité avec le caoutchouc naturel. C'était sans doute une de ces recherches que l'on effectue sans grand espoir, car les produits que donnent les autres métaux alcalins frères du lithium étaient déjà bien connus comme de pâles imitations du caoutchouc naturel. Cette découverte ne semble pas moins tardive, et l'on doit se demander pourquoi les investigations n'ont pas été menées simultanément avec un zèle égal parmi tous ces groupes de métaux.

C'est que le nombre des chercheurs qui fréquentent les laboratoires scientifiques et industriels est bien petit, en face de la multiplicité des sujets de recherche. *Pour pallier cette faiblesse et affronter la concurrence des pays plus favorisés sous ce rapport, un mouvement d'« action concertée » se développe actuellement en France. Des réunions de savants dressent l'inventaire des problèmes dont l'urgence est manifeste, s'efforcent de coordonner les bonnes volontés et d'obtenir les crédits indispensables.*

Le polyisoprène 1-4 cis triomphera-t-il du produit naturel? Il serait téméraire de l'affirmer, car, à une baisse du prix de revient de l'isoprène répondrait peut-être une amélioration du rendement des plantations d'hévéas.

Le succès de la synthèse d'un caoutchouc identique à celui que la nature nous offre ne doit pas faire oublier l'existence de succédanés nombreux et variés plus largement utilisés que le caoutchouc lui-même.

On verra, en effet, à propos des propriétés physiques des hauts polymères, que les structures chimiques qui entraînent une grande élasticité satisfont à des caractères très généraux.

Aussi, parmi les succédanés du caoutchouc, ne figurent pas seulement des substances qui lui sont voisines, comme les polybutadiènes et les polychloroprènes, mais aussi beaucoup d'autres, plus lointaines. Citons le G. R. S., copolymère de butadiène et de styrène, le polybutadiène 1-4 cis, ainsi que les copolymères d'éthylène et de propène, promis à un grand avenir en raison de leur très bas prix.

comment distinguer le caoutchouc des substances caoutchouteuses de synthèse ?

Nous avons affirmé que les substances caoutchouteuses préparées respectivement en partant de l'isoprène et du butadiène différaient par leurs propriétés physiques et leurs structures. Il est facile de distinguer entre ces structures lorsque l'on compare l'absorption des radiations infrarouges par divers échantillons.

Quittons le laboratoire de chimie, souvent encombré et malodorant, pour pénétrer dans la pièce, méticuleusement ordonnée, où siège le spectrographe infrarouge. Comme celui de tous les appareils de physique commerciaux, son aspect est simple; le profane remarque surtout un grand cylindre horizontal, sur lequel une feuille de papier portant des graduations est enroulée et qu'un stylet rempli d'encre rouge s'apprête à parcourir. Le produit est généralement présenté sous la forme d'un film très mince obtenu par évaporation de quelques gouttes de sa solution benzénique étalées sur une surface de sel gemme (chlorure de sodium). L'échantillon, avec son support, est placé sur le trajet du faisceau incident de rayons infrarouges.

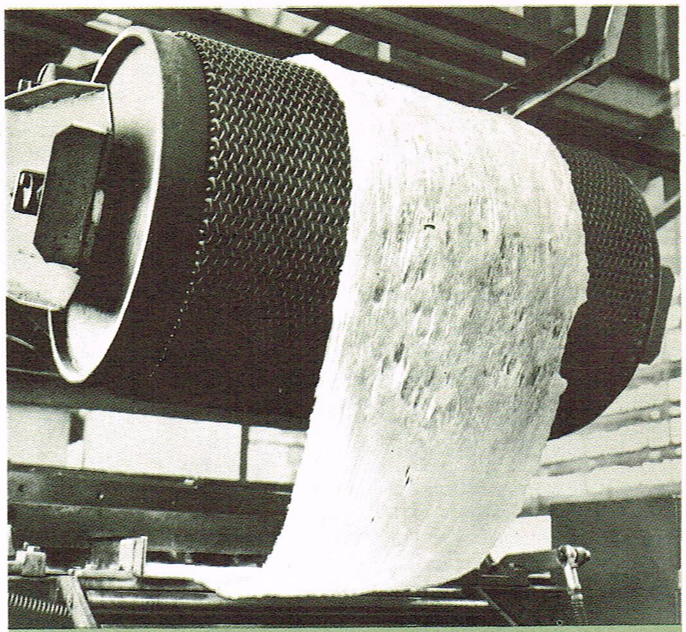
Le cylindre enregistreur tourne, et, lentement, s'inscrit une courbe aux multiples accidents.

L'examen des spectres porte avant tout sur la position des maximums d'absorption sur l'échelle des longueurs d'onde. Ils se signalent par des pics ou de simples épaulements, dont la liste constitue une sorte de fiche signalétique du produit. Judicieusement commentée, elle permet d'acquérir des précisions sur la structure de ses molécules et même de doser la proportion de chacun des « motifs ». Les maximums les plus caractéristiques de l'isoprène sont rassemblés ci-dessous :

1-2	3-4	trans 1-4	cis 1-4
10,97 μ	11,24 μ	8,65 μ	8,85 μ

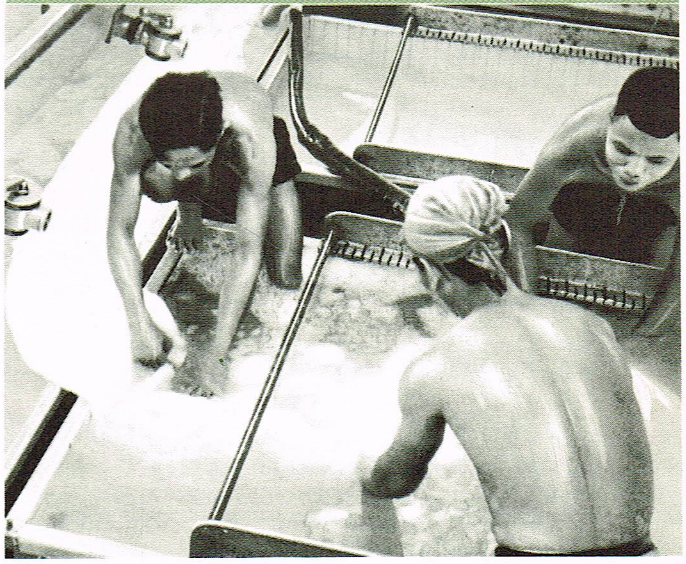
●

Lorsqu'un morceau de lithium, ou d'un autre métal alcalin fraîchement découpé, est plongé dans de l'isoprène, il se recouvre peu à peu d'une couche gélatineuse de caoutchouc imbibée de monomère. Longtemps, l'opinion a prévalu que la polymérisation s'effectuait sur la surface du métal; ne cessait-elle pas si le métal était séparé du liquide? Les recherches les plus récentes ont infirmé cette conclusion de façon décisive. Si l'on utilise comme promoteur du lithium métallique très divisé, le liquide reste actif même après passage à travers une membrane millipore, qui retient les particules solides; bien plus, ce filtrat provoque la polymérisation du styrène que l'on vient à lui ajouter, et les macromolécules finales sont constituées par une séquence de motifs 1-4 cis dérivés de l'isoprène entre deux séquences de motifs styrène.



La bande de caoutchouc butyl est refroidie avant le conditionnement.

La bande de caoutchouc naturel sort des blocs de coagulation.



COMMENT LA NATURE RÉALISE LA SYNTHÈSE DU CAOUTCHOUC

Plus d'un millier d'espèces végétales produisent du caoutchouc : grands arbres (hévéa ou ficus), arbustes (guayule, lianes) et herbacées (kok-saghyz, variété de pissenlit qui connut en Russie son heure de gloire). Mais par l'intensité et la qualité de sa production, la facilité de l'extraction de son caoutchouc, c'est *Hevea brasiliensis* qui, à l'heure actuelle, domine encore tous ses concurrents. Si l'on pratique sur son tronc une saignée transversale, un liquide abondant, d'un blanc laiteux, s'écoule : le latex.

Il contient en suspension 30 à 40 p. 100 de fines particules solides, auxquelles le caoutchouc proprement dit participe pour près de 90 p. 100. Le rendement, qui varie avec la nature des hévéas, atteint chaque année 1 000 à 6 000 litres de latex par hectare, c'est-à-dire pour 250 à 300 arbres; cela représente 300 à 2 400 kg de caoutchouc.

Si l'hévéa blessé perd son latex comme un animal perd son sang, c'est que, comme le sang, le latex est contenu dans un système de vaisseaux communiquant entre eux par de multiples ramifications. L'analogie s'arrête là; les arbres ne possèdent pas de système circulatoire actionné par une pompe comparable au cœur des mammifères, et les canaux laticifères se vident par le seul effet de la pesanteur.

le latex, matière vivante

Un examen minutieux de ces canaux, sous les grossissements extrêmes du microscope électronique, montre leur paroi cellulosique tapissée d'une fine membrane. Le latex qui les remplit entièrement n'est pas logé, comme on s'y attendrait, dans des

Phot. Esso, Institut fr. du caoutchouc.

vacuoles rompant la continuité d'un cytoplasme. C'est lui-même un cytoplasme hautement différencié, une matière vivante, et non un simple produit d'excrétion; ne contient-il pas des noyaux cellulaires et n'est-il pas capable de réaliser la synthèse du caoutchouc?

Jusqu'ici, nous n'avons guère distingué dans le latex autre chose que des globules de caoutchouc dispersés dans un liquide aqueux. C'est là une description sommaire; ces globules, d'un diamètre allant de 0,15 à 1,5 μ, sont recouverts d'une couche fine de protéines, qui les maintient en suspension.

Ils voisinent avec d'autres particules de nature différente, beaucoup moins nombreuses, et l'eau qui les baigne recèle du phosphate de potassium, des sucres, des acides ribonucléiques, des pyrophosphates de nucléosides, des polypeptides sulfurés, etc. Donner force détails sur ces combinaisons serait fastidieux; notons seulement qu'une partie d'entre elles communique au latex une activité synthétique remarquable.

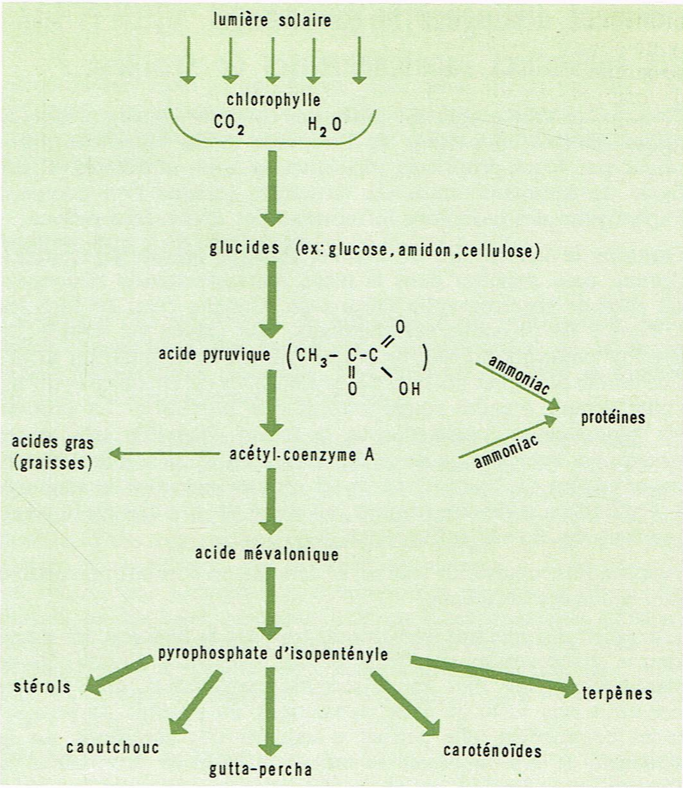
Pour étonnante qu'elle soit, cette activité ne reflète que partiellement celle de la plante entière, qui réalise la synthèse du caoutchouc à partir de matériaux minéraux aussi simples que l'anhydride carbonique et l'eau. Une telle transformation ne peut se produire sans de multiples étapes intermédiaires.

le mécanisme d'une synthèse naturelle

On est amené à ouvrir une large enquête sur la nature des composés chimiques pouvant servir à l'élaboration du caoutchouc.

Pour effectuer ces études, on a alimenté la plante soit par cette voie naturelle que constituent les racines, soit par l'immersion des tiges sectionnées, soit encore par badigeonnage des feuilles avec des liquides chargés en substances nutritives, elles-mêmes marquées par des atomes radio-actifs, afin qu'on pût suivre leurs transformations.

Progressivement, les détails de celles-ci se sont révélés, et l'on a reconnu que la voie menant au caoutchouc bifurque aussi vers les terpènes. Parmi eux, citons le citronellal (essence de citronnelle), l'α-pinène (essence de térébenthine), le camphre; on peut aussi rattacher à ce groupe les carotènes (colorants de la carotte)



La place de la synthèse du caoutchouc dans l'activité chimique de la cellule vivante.

et même le cholestérol, objet fameux d'inquiétudes pour la santé des artères.

Cette origine commune avait d'ailleurs été pressentie, car on savait depuis longtemps que les chaînes carbonées des molécules de ces produits peuvent se construire, au moins sur le papier, à partir de un à huit chaînons isopréniques.

LES ÉTONNANTES PROPRIÉTÉS PHYSIQUES DES COMPOSÉS MACROMOLÉCULAIRES

Les substances macromoléculaires possèdent des propriétés particulières, qui se manifestent déjà en solution, mais de façon plus directement sensible à l'état solide; ce sont elles qui en justifient l'emploi pour la fabrication de toutes sortes d'objets.

Certaines macromolécules se rapprochent des métaux; comme eux, elles se laissent extruder, après fusion, en fils ou en tubes, laminer en feuilles, couler dans des moules, et peuvent supporter des efforts considérables à l'extension ou à la compression; fragiles aux chocs parfois, elles allient souvent à la solidité une malléabilité qui leur évite le sort du verre dans nombre d'applications.

D'autres se dissolvent dans certains solvants organiques, leur mise en forme étant alors obtenue à partir de solutions concentrées appelées « collodions ». Pour celles-ci, il faut évidemment éviter de les mouiller ensuite avec leurs solvants. Ainsi, une goutte de vernis à ongles peut, si on n'y prend pas garde, provoquer des trous dans certains tissus.

D'autres encore sont insolubles dans tous les liquides, à moins qu'une attaque d'ordre chimique ne les transforme, ou ne les détruise.

Il est enfin une propriété tout à fait curieuse, la haute élasticité acquise par le caoutchouc lorsque celui-ci a été légèrement vulcanisé. Il suffit de tirer dessus pour en voir la longueur multipliée par plus de 10, puis il revient à son état initial lorsque l'effort cesse. *De plus, il cristallise lorsqu'on l'étire* (ce que révèle l'observation de la diffraction des rayons X par l'échantillon).

Bien d'autres composés macromoléculaires se comportent de façon identique au caoutchouc. L'un d'entre eux même, le polychlorure de phosphonitrile, est purement minéral, puisqu'il ne contient pas de carbone, mais seulement du phosphore, de l'azote et du chlore.

Les solutions de hauts polymères présentent également des caractères particuliers. Le processus de la dissolution lui-même ne ressemble pas à la disparition d'un morceau de sucre dans un verre d'eau; un fragment de caoutchouc ou de polystyrène plongé dans du benzène commence par augmenter de volume en s'imprégnant d'une proportion considérable du liquide; c'est alors seulement qu'il commence à disparaître. Les solutions sont visqueuses, diffusent la lumière; dans une centrifugeuse rapide, elles se concentrent vers le fond du godet; placées dans un sac de collodion, elles laissent exsuder le solvant pur.

Une des préoccupations dominantes des chercheurs est d'expliquer toutes ces propriétés par les caractères, généraux ou particuliers, de la structure des macromolécules. C'est la transposition à ce domaine de soucis millénaires; en effet, la théorie atomique elle-

même a probablement été inventée, plusieurs siècles avant J.-C., pour donner une image intuitive du mélange de deux substances apparemment continues. Bien plus tard, c'est toujours à partir du comportement des particules ultimes que Descartes chercha à comprendre celui de la matière qui tombe sous nos sens. « ... vous pouvez imaginer même différence entre de l'eau et de la glace, que vous feriez entre un tas de petites anguilles, soit vives, soit mortes, flottantes dans un bateau de pêcheur tout plein de trous par lesquels passe l'eau d'une rivière qui les agite, et un tas des mêmes anguilles, toutes sèches et raides de froid sur le rivage. »

En France, l'étude des relations entre structure et propriétés constitue l'un des buts principaux du Centre de recherches sur les macromolécules, dirigé par M. Ch. Sadron.

COMMENT LA RÉTICULATION REND INSOLUBLES LES MACROMOLÉCULES

Ne nous attardons pas à discuter pour quelles raisons le polystyrène est soluble dans le benzène et insoluble dans l'alcool; on ne possède que peu de lumière sur les questions de ce genre; malgré des tentatives modernes pleines d'intérêt, nous nous contenterons du vieil adage « les semblables se dissolvent », et, en effet, le polystyrène porte de multiples groupes benzéniques (semblables au benzène lui-même). Observons, cependant, que la solubilité de cette macromolécule devient de plus en plus faible lorsque sa masse moléculaire augmente. Cette solubilité s'annule même complètement pour des dimensions accessibles au microscope, disons de l'ordre du micron (0,001 mm).

Si l'on désire des polymères insolubles dans quelque solvant que ce soit, ce qui peut être bien utile, il faut donc fabriquer des molécules infinies. **Le meilleur moyen pour y parvenir est de « réticuler » les macromolécules de taille moyenne dont la réunion constituait un fragment de substance soluble, c'est-à-dire de les relier entre elles par de nouvelles et courtes chaînes.**

Pour cette raison, la gélatine qui recouvre les plaques et les films photographiques a été traitée par le formol, sinon elle s'évanouirait dans le révélateur et le fixateur, et au cours du lavage. Les nombreuses résines, qui, mélangées à des charges et à divers additifs, puis moulées à chaud, se transforment en solides infusibles, subissent des réticulations; l'objet terminé peut être considéré comme une seule et unique molécule.

Un exemple méritant d'amples développements est celui de la vulcanisation du caoutchouc, dont les inventeurs furent Goodyear (1839) et Hancock (1842). Cette opération consiste à chauffer le caoutchouc mélangé à du soufre, à une température allant de 120 à 150 °C; elle dure plusieurs heures, mais l'addition de certains composés, notamment de disulfure de tétraéthylthiurame, à la dose de 1 p. 100, la réduit à quelques minutes.

Le soufre ne peut plus être extrait; il est combiné aux chaînes macromoléculaires, établissant ainsi des « ponts » entre elles.

Si ces liens sont peu nombreux (2 à 4 p. 100 de soufre), le produit ne se dissout déjà plus dans le benzène, mais, après un contact prolongé avec ce liquide, il se gonfle énormément et devient une gelée. Par ailleurs, le produit sec a acquis une haute élasticité, sur laquelle nous reviendrons ultérieurement.

Lorsque le nombre des liens devient grand, résultat auquel on arrive en augmentant la proportion de soufre jusqu'à 35 à 50 p. 100, le gonflement intempestif disparaît; c'est l'ébonite, dure et cassante, utilisée longtemps, notamment à la fabrication des stylos.

les résines polyesters

Les applications de la réticulation sont multiples; les peintures à l'huile de lin ne sèchent pas, comme on le dit couramment, ou plutôt l'évaporation des constituants volatils n'est qu'un épisode assez secondaire; le processus essentiel consiste en la polymérisation de l'huile de lin et la réticulation des macromolécules formées, sous l'influence de l'oxygène de l'air et de divers additifs.

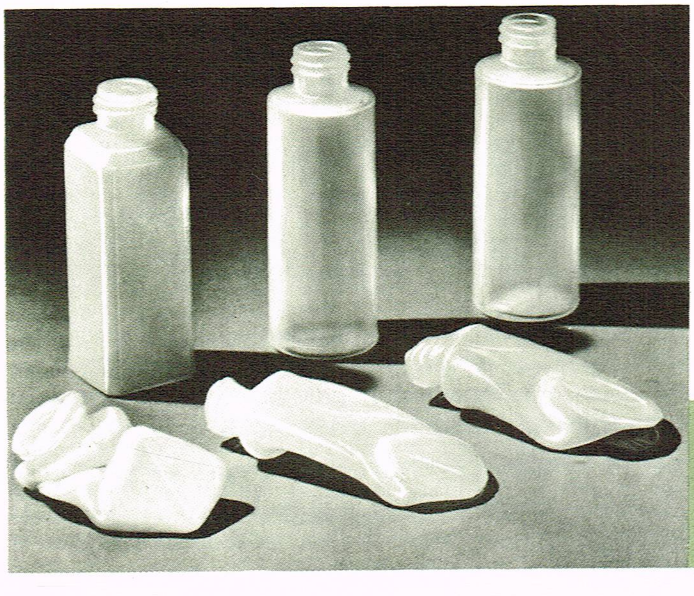
Plus modernes sont les résines polyesters, dont l'industrie, celle des vernis et des peintures en particulier, fait une large application depuis ces dernières années. Ce sont elles aussi qui forment la matière de ces plateaux décorés dont tous les magasins regorgent. Le composé macromoléculaire de base est un polyester, le polymaléate d'éthylène-glycol. On l'obtient en faisant réagir deux produits peu coûteux, l'anhydride maléique et l'éthylène-glycol. Ce sont maintenant des chaînes de polystyrène qui vont lier entre elles les macromolécules du polyester; il suffit pour cela de le dissoudre dans du styrène monomère, et de provoquer la polymérisation du mélange par un promoteur tel que le peroxyde de benzoyle. Des modifications de détail apportées au schéma précédent ont pour but d'en rendre la réalisation commode.

la réticulation par liaisons hydrogène explique les propriétés des polyamides

Souvent les composés macromoléculaires doivent leur insolubilité à une sorte de réticulation spontanée. Ainsi les polypeptides, et plus généralement un grand nombre de polyamides, parmi lesquels ceux qui portent les marques Nylon et Rilsan, sont constitués par des macromolécules disposées parallèlement en nappes, du fait de liaisons latérales dites « liaisons hydrogène ».

Nous les avons figurées en pointillé sur les formules correspondantes; on remarquera qu'un atome d'hydrogène lié à un atome d'azote se trouve « capté » par un oxygène d'une macromolécule voisine.

C'est d'ailleurs pour une raison semblable que l'eau devient vapeur à 100 °C seulement, au lieu d'être un gaz comme l'hydrogène sulfuré. Alors que les molécules H_2S n'ont aucune tendance à se lier, deux ou plusieurs molécules d'eau forment des agrégats sous la contrainte des liaisons hydrogène.



L'exposition des matières plastiques aux radiations ionisantes peut provoquer une réticulation qui accroît leur résistance à la chaleur. Des six flacons stérilisés à une même température, trois subissent un ramollissement qui les rend inutilisables, alors que les trois autres, préalablement irradiés, ne subissent aucun dommage.

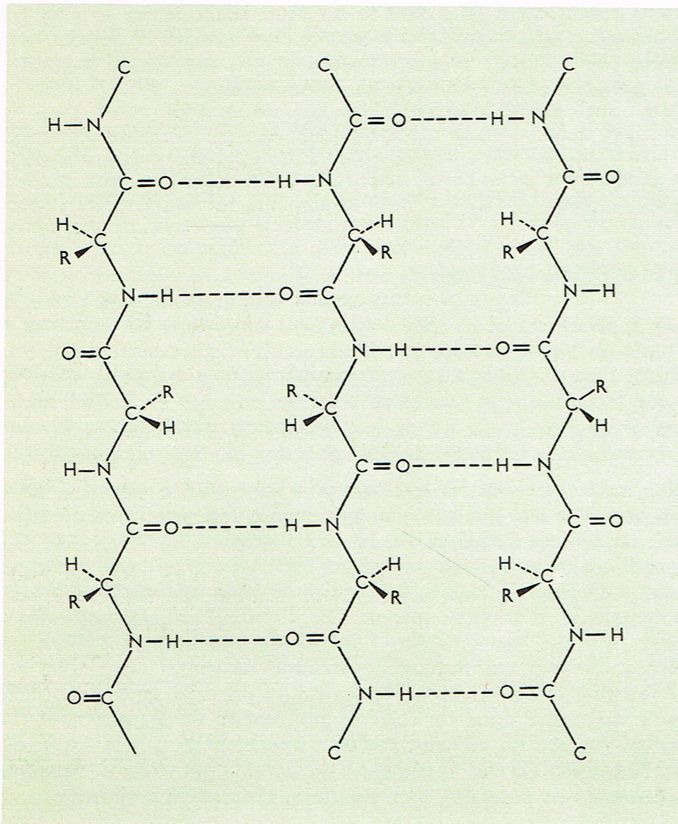
Phot. U. S. I. S.

Liaisons dont l'importance immense s'est dégagée au cours des vingt-cinq dernières années ; c'est à elles que nous devons la solidité des fils de soie ou de Nylon, et même des cheveux, aussi bien que la dureté des ongles et le rassissement du pain.

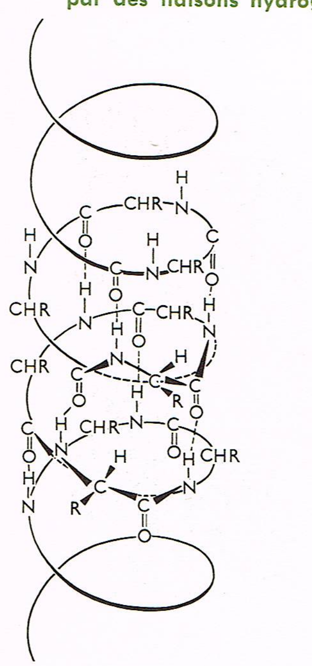
Cependant cette « liaison hydrogène » se rompt avec beaucoup moins d'efforts que les valences normales; c'est pourquoi les nappes de polyamides se désagrègent et se dissolvent dans des liquides tels que l'acide formique, le phénol, l'hydrate de chloral, car ils contiennent des atomes d'hydrogène capables de se substituer dans leur rôle à ceux que portent les azotes du polyamide.

L'usage de ces produits est, heureusement pour nos vêtements, absolument inhabituel dans la vie quotidienne.

POURQUOI LE CAOUTCHOUC EST-IL SI ÉLASTIQUE ?



Trois chaînes polypeptidiques liées entre elles par des liaisons hydrogène, figurées en pointillé.



Les macromolécules s'enroulent volontiers en hélice. En particulier, les polypeptides (protéines simplifiées) peuvent affecter cette forme, concurremment avec celle en zigzag. La stabilité de l'édifice est encore assurée par des liaisons hydrogène qui soudent par points une spire à l'autre.

La haute élasticité manifestée par le caoutchouc modérément vulcanisé, c'est-à-dire réticulé, mérite quelques éclaircissements. Pour être en mesure de les donner, il est nécessaire de considérer d'assez près l'aspect que prend une macromolécule lorsqu'elle se dissout. Nous avons vu que sa longueur atteint 1 000 à 10 000 fois son diamètre, ou plus encore; malgré cela, il en est de rigides; en solution, elles restent de véritables bâtonnets. Tel est le cas de beaucoup de polypeptides naturels et synthétiques; l'enroulement en hélice de la chaîne principale, que nous avions effectué par la pensée pour comprendre la nature des molécules isotactiques, s'est produit alors de façon effective par le jeu des forces naturelles. L'hélice se trouve verrouillée par les liaisons hydrogène qui s'établissent d'une spire à l'autre. Comme celles qui ont été décrites plus haut, elles unissent un groupe $> N - H$ à un groupe $> C = O$, mais cette fois à l'intérieur d'une même macromolécule.

En général, de telles forces de liaison sont absentes; heurtée perpétuellement par les petites molécules du solvant, la macromolécule subit en un temps très court toutes les déformations possibles. On se fera une idée approximative de l'aspect qu'elle prend en considérant un modèle simplifié à l'extrême, la chaîne d'arpenteur librement articulée, lorsqu'on la replie de façon totalement désordonnée. Pour plus de rigueur, il faudrait tenir compte de ce que les déformations permises à la chaîne macromoléculaire ne sont pas aussi libres, mais limitées à des rotations autour des liaisons simples.

Finalement, nous sommes en présence d'une espèce de pelote enchevêtrée et lâche, au travers de laquelle le solvant circule. Sa complication même, bien loin d'interdire tout espoir d'acquérir quelques données précises sur ses caractéristiques, permet au contraire, puisqu'elle est soumise aux lois des grands nombres, de lui appliquer les méthodes du calcul des probabilités. En premier lieu, on détermine la valeur moyenne, $\overline{r^2}$, du carré de la distance qui sépare les deux extrémités de la pelote.

Pour le modèle simplifié de la chaîne d'arpenteur à articulations libres, on démontre la relation (loi des grandes séries; cf. au chapitre « la Mesure de la connaissance ») au tome II.

$$\overline{r^2} = n\ l^2,$$

n étant le nombre de segments de la chaîne, supposés tous de même longueur l.

Supposons maintenant que, par un moyen quelconque, nous tirions sur les extrémités d'une macromolécule, de façon à lui donner une forme plus étirée. Il est nécessaire pour y parvenir d'exercer un effort, puisqu'elle tend à reprendre sa configuration moyenne sous l'effet de l'agitation moléculaire. Si la traction cesse, elle se raccourcit et revient effectivement à la configuration primitive.

En poussant jusqu'au bout ce raisonnement, on voit que l'allongement maximal d'une chaîne d'arpenteur serait $n\ l$, soit pratiquement de $\sqrt{n}$ fois (100 fois pour une chaîne composée de 10 000 maillons) la dimension initiale $\sqrt{n}l$ de la macromolécule.

Telle est l'origine profonde du phénomène de haute élasticité. Pour préciser encore mieux sa nature, passons des macromolécules en solution aux mêmes particules condensées à l'état liquide. Les calculs statistiques qui avaient été effectués lorsqu'elles étaient éloignées les unes des autres, tout insuffisants qu'ils

soient maintenant, donnent une approximation de la réalité. Si la température du liquide baisse, il se solidifie, mais les macromolécules de caoutchouc gardent une mobilité relative jusqu'à — 70 °C (203 °K). Au-dessus, le solide doit plutôt être considéré comme un liquide très visqueux, et la configuration moyenne de ses macromolécules reste à peu près la même qu'en solution. Tirer sur les deux extrémités d'un fil aurait apparemment pour effet d'étirer les molécules; cela n'est vrai que pour une part très modeste, car elles glissent les unes sur les autres, ce qui aboutit à une déformation permanente, sans retour spontané à l'état initial. Pour les empêcher de glisser, il suffit de les attacher les unes aux autres de loin en loin, c'est-à-dire de les réticuler.

Du même coup, il devient clair que l'élasticité n'appartient pas exclusivement au caoutchouc. **N'importe quelle substance macromoléculaire est élastique sous la double condition qu'elle soit réticulée, fût-ce modérément, et que ses chaînons à l'état solide jouissent d'une relative mobilité.** Si le nombre des liens augmente, toute déformation devient impossible; le caoutchouc se transforme en ébonite.

Nous venons d'ouvrir de vastes horizons industriels. Pourquoi ne pas chercher alors à fabriquer des substances caoutchouteuses toutes différentes du polyisoprène? On n'y a pas manqué; des produits remarquables par certaines qualités particulières sont depuis longtemps lancés sur le marché. Les efforts portent actuellement sur l'utilisation de matières premières peu coûteuses; le polybutadiène 1-4 cis, les copolymères de l'éthylène et du propène semblent voués à un grand avenir; une difficulté qui semble avoir été résolue maintenant est celle de la vulcanisation (réticulation) de ces derniers. Le meilleur moyen consiste à ajouter au mélange d'éthylène et de propène une certaine proportion de butadiène ou de cyclopentadiène; ces derniers s'intègrent dans les chaînes macromoléculaires, les rendant attaquables par les peroxydes ou par le soufre. Ces nouveaux matériaux, en dehors de leur bas prix, ont l'avantage d'être peu sensibles à l'oxygène de l'air, car ils contiennent peu ou pas de liaisons éthyléniques.

Plus anciens, les polyisobutènes, ou caoutchoucs butyles, très peu perméables au gaz, assurent l'étanchéité des chambres à air. Ce sont des copolymères de l'isobutène et du butadiène, ce dernier étant introduit, comme dans l'exemple précédent, pour permettre une vulcanisation facile.

un fil cristallise lorsqu'on l'étire !

Lorsqu'on étire un fil de caoutchouc, une modification de sa structure à l'échelle moléculaire se produit simultanément : de l'état amorphe, il passe à l'état cristallisé. Ne nous attendons pas à voir des cristaux comparables, par leurs formes tranchées et leur grosseur, à ceux du sel de cuisine; il s'agit de minuscules cristallites noyées dans une gangue amorphe, que la diffraction des rayons X révèle.

Le phénomène de cristallisation par étirement n'apparaît pas seulement avec le caoutchouc et ses succédanés, mais aussi avec les fibres synthétiques à base de polyamides, de polyéthylène basse pression et de polypropène isotactique; ces fibres, peu élastiques, conservent à la fois leur nouvelle longueur et leur état cristallin. N'oublions pas que celui-ci, plus compact qu'à l'état amorphe, est le siège de forces d'attraction maximales entre des chaînes étendues parallèlement.

Des modifications sensibles de propriétés physico-chimiques en résultent; dans ce nouvel état, la substance ne se laisse plus teindre par les colorants qu'elle absorbait auparavant, et sa solubilité générale a diminué. Dans certains cas, les cristallites ne survivent pas à la traction qui les a engendrées.

Une expérience curieuse et convaincante consiste à suspendre à l'extrémité d'un fil d'alcool polyvinylique un poids juste suffisant pour lui appliquer une tension, et insuffisant pour provoquer son allongement. Si l'on plonge alors ce fil dans de l'eau bouillante, tant qu'on maintient le poids au-dessus du fond du récipient il ne se passe rien. Si l'on descend progressivement le fil jusqu'à ce que le poids atteigne le fond du récipient, la tension qu'il exerce se relâche alors et le fil se dissout à ce moment (et à ce moment seulement) d'une façon instantanée.

quelques techniques de finissage des fils : moulinage et texturation

Les fils que forment les produits cités plus haut sont au contraire stables dans leur nouvelle microstructure, et plus solides bien que plus fins. Aussi, dès que les brins qui les constituent se sont réunis au sortir des trous minuscules de la filière, on leur fait subir une traction qui multiplie leur longueur par 4 à 5. Mais ils ne sont pas encore prêts pour le tissage. Leur présentation en brins parallèles est en général peu favorable à cette opération. Le moulinage, qui consiste à tordre le fil sur lui-même de quelques dizaines, voire quelques centaines de tours au mètre, augmente sa cohésion.

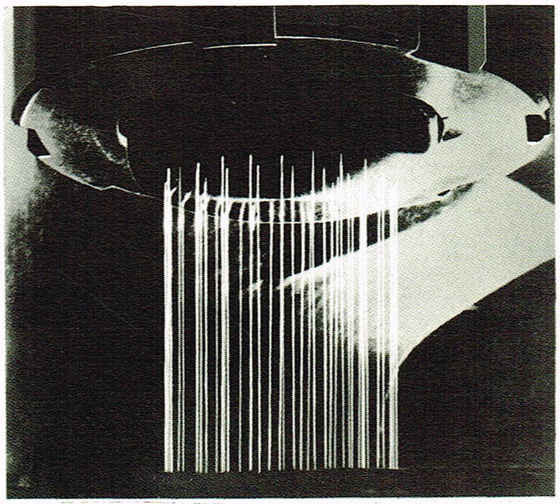
Ce procédé classique diminue cependant la voluminosité du fil, c'est-à-dire le volume qu'il occupe dans l'espace. De nouveaux procédés, dits de « texturation », permettent au contraire d'augmenter cette voluminosité tout en conservant une bonne cohésion.

Les brins individuels sont frisés, ce qui les écarte les uns des autres et leur confère un gonflant extrêmement flatteur. Dans le procédé « Hélanca », le fil est tordu de quelques milliers de tours par mètre, étuvé pour fixer cette torsion, puis détordu complètement, en dépassant même de quelques tours la torsion zéro.

Le procédé dit de « fausse torsion », qui paraît l'un des plus économiques, consiste à faire glisser le fil sur un cylindre tournant à grande vitesse, cela parallèlement à son axe.

Dans le procédé utilisé pour les articles « Banlon », la frisure est obtenue par compression du fil entre deux rouleaux, puis dans une chambre chauffée à 200 °C, où la frisure est fixée. Cette fixation thermique devient inutile dans le procédé « Taslan », où la séparation des brins est obtenue par un violent courant d'air.

L'aspect de ces fils « texturés », leur toucher et leurs propriétés d'isolation thermique se trouvent considérablement améliorés. Le filage est donc une opération industrielle délicate, aux multiples aspects, dont le secret se justifie par l'ampleur des travaux nécessités par sa mise au point.

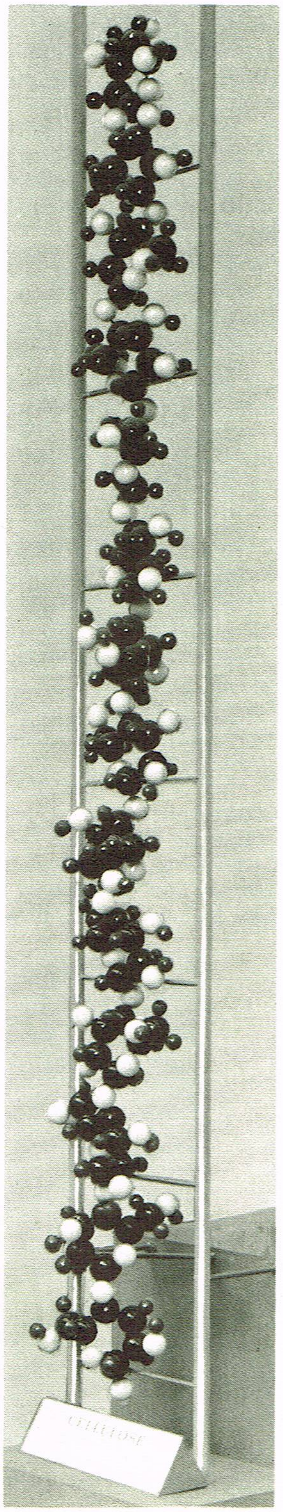


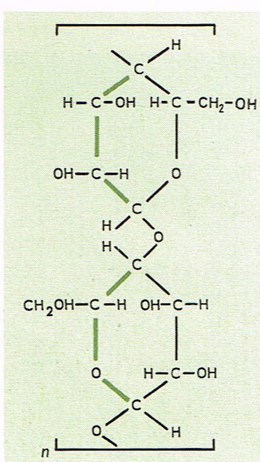
Brins sortant de la filière dans une opération de de filature à sec.

Phot. A. Gamet.

LES MACROMOLÉCULES DE LA VIE

Les pages qui précèdent ont donné un aperçu de la variété des espèces macromoléculaires dont la synthèse a été réalisée au laboratoire, et des possibilités grandioses qui s'offrent encore aux chimistes. A côté d'eux, la cellule vivante semble faire bien piètre figure, puisque les macromolécules qu'elle fabrique rentrent dans quatre groupes seulement, très nettement tranchés : polyisoprènes, polysaccharides, protéines, acides nucléiques. Et encore, seuls les deux derniers, communs à l'ensemble des organismes connus, peuvent être vraiment considérés comme inséparables de la vie.





Motif élémentaire de la cellulose, répété dans la macromolécule construite avec des boules de couleurs et de dimensions conventionnelles.

Les polyisoprènes, composés de carbone et d'hydrogène, ont déjà été décrits précédemment. Leur fonction dans la plante demeure mystérieuse ; on ne saurait même exclure qu'ils n'en aient aucune et qu'on se trouve devant un mécanisme vital fonctionnant sans but.

LES POLYSACCHARIDES

Nous connaissons bien la cellulose, l'amidon, le glycogène, les gommes, les mucilages. Papier, toile, pain, gelées de confiture, ce sont là leurs applications familières, qui rappellent le rôle de soutien des tissus végétaux ou de réserve alimentaire que ces substances jouent dans la nature.

Une autre variété, la chitine, sorte de cellulose aminée, entre dans la structure d'animaux inférieurs. La carapace des crustacés, les élytres des insectes contiennent une forte proportion de chitine, durcie par des dépôts de carbonate de calcium. Les chondroïtine-sulfates des cartilages, l'acide hyaluronique des tissus conjonctifs, l'héparine, puissant anticoagulant naturel du sang, les polysaccharides des érythrocites, aux subtiles propriétés antigéniques, se rapprochent de la chitine.

Ainsi les polysaccharides apparaissent comme des figurants actifs, mais souvent un peu secondaires, et l'on conçoit que certains êtres s'en passent presque totalement.

Leurs molécules ont une structure relativement simple, puisqu'elles résultent de la répétition d'un seul motif, souvent voisin du glucose.

Dans l'amidon et le glycogène, ce motif élémentaire dérive directement de la forme α du glucose, et, dans la cellulose, de la forme β.

Que la cellulose soit fournie par le coton, la ramie, le bois ou les bactéries, elle se présente toujours sous forme de fibres, parce que ses macromolécules sont elles-mêmes de longs fils comprenant plusieurs milliers de motifs élémentaires enchaînés linéairement les uns aux autres.

Au contraire, l'amidon et le glycogène, bien que leur motif élémentaire diffère peu de celui de la cellulose, sont des poudres et ne sauraient être étirés en fibres; à cette morphologie, il faut rattacher un fait, mis en lumière grâce à des études chimiques variées : leurs macromolécules sont ramifiées.

LES INDISPENSABLES PROTÉINES

Le groupe des protéines comporte un nombre considérable de représentants, dont certains sont très communs : le blanc d'œuf, la caséine du lait, le caillot formé par le sérum sanguin, les cheveux, les ongles, la laine, la soie.

Beaucoup de protéines jouent un rôle très particulier, propre à chacune. Pepsine, trypsine, papaïne, amylase sont des termes qui recouvrent certaines familles d'enzymes hydrolysants. D'autres familles sont impliquées dans des phénomènes d'oxydation et de réduction; elles contiennent alors un coenzyme, de nature chimique très différente. Il en est de même de l'hémoglobine, formée d'une partie protéique, la globine, et du dérivé ferrique de la protoporphyrine.

L'activité biochimique ou physiologique des protéines fournit un moyen sensible de les suivre au cours des tentatives d'extraction, puis de purification, si bien que nombre d'entre elles ont été obtenues à l'état cristallisé, et sans doute assez pures.

Deux protéines apparemment identiques par leurs propriétés physiques et de même fonction biologique, mais provenant d'êtres vivants d'espèces distinctes, présentent de subtiles différences de composition, dont nous reparlerons, et se distinguent par leurs propriétés physiologiques; injectée à un animal, chacune provoque la formation d' « anticorps » spécifiques, donnant un précipité avec le même échantillon de protéine, et non avec d'autres.

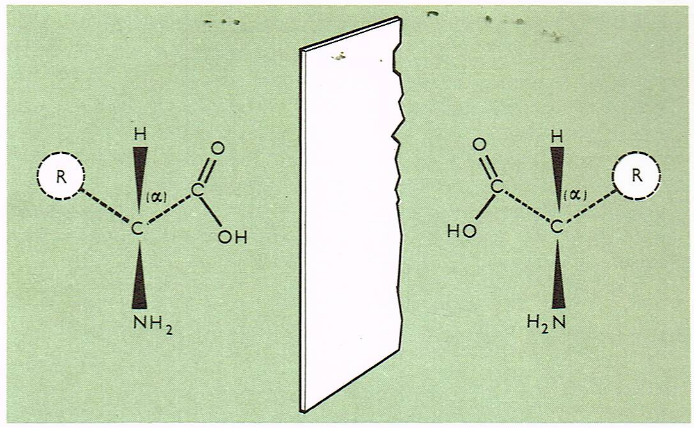
LES ACIDES AMINÉS, MATÉRIAUX COMMUNS A TOUTES LES PROTÉINES

Sous leur diversité, elles possèdent un trait commun : ce sont des polyamides; et elles appartiennent à ce titre à la famille du Nylon et du Rilsan. On peut, en effet, les considérer comme le résultat de la polycondensation d'acides α-aminés, avec élimination d'eau.

La liste des acides α-aminés biologiques groupe vingt et un représentants, répondant tous à la formule générale ci-contre, dans laquelle la fonction amine est immédiatement voisine de la fonction acide (en position α). C'est le radical R qui varie d'un type d'acide α-aminé à un autre.

Chacun des acides α-aminés, sauf le plus simple, le glycocolle, comporte deux variétés dites « énantiomorphes », différant uniquement par la disposition dans l'espace des quatre radicaux portés sur l'atome de carbone (α). On vérifiera sur la figure que ces deux dispositions, symétriques par rapport au plan P, ne sont pas identiques, c'est-à-dire qu'elles ne peuvent être superposées par un déplacement dans l'espace, pas plus que la main droite ne peut être superposée à la main gauche.

Les acides α-aminés forment donc deux séries, désignées par L et D; les produits naturels sont tous de la série L.



Formule générale des acides α-aminés. De l'un à l'autre, R diffère. Ils se partagent en deux séries, symétriques par rapport au plan P, et, en conséquence, non superposables.

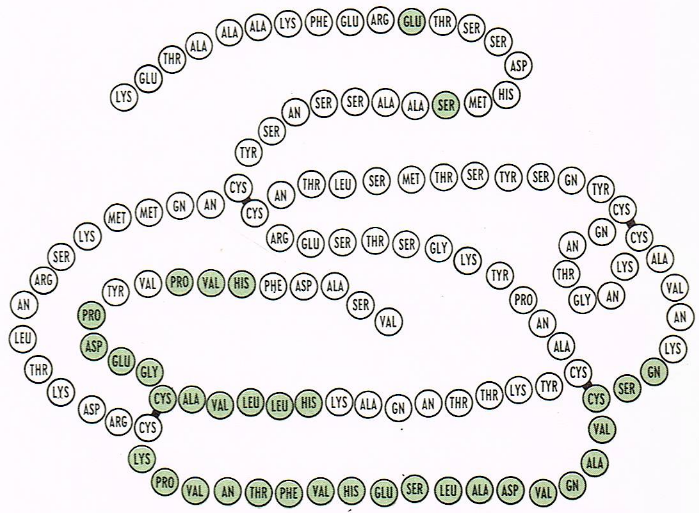
●

Souvent, les vingt et un acides aminés entrent tous dans la constitution d'un même échantillon de protéine, avec des pourcentages relatifs très divers. Un problème se pose immédiatement : les motifs d'acides aminés se succèdent-ils dans un ordre variable d'une macromolécule à l'autre, ou bien, au contraire, dans un ordre rigoureusement déterminé? On pariera sans témérité pour la seconde hypothèse.

L'évolution et la reproduction d'un individu, suivant un schéma inéluctable, toujours identique de génération en génération, ne sauraient s'accommoder de matériaux hétéroclites.

Plusieurs équipes de biochimistes se sont attachées à déterminer selon quelles suites précises s'ordonnaient les acides aminés, en commençant par les espèces protéiques de masse moléculaire faible, telle la ribonucléase, extraite du pancréas de bœuf. Cet enzyme provoque l'hydrolyse de l'acide ribonucléique (A. R. N.) en fragments plus petits.

Ses 124 aminoacides sont indiqués ci-contre, dans l'ordre de leur enchaînement. On remarquera une complication dont nous n'avons pas encore parlé : des liaisons sont établies à l'intérieur de la chaîne par quatre paires d'atomes de soufre, si bien que celle-ci ne peut être qualifiée de linéaire, mais de polycyclique.



La structure de la ribonucléase. Les traits figurent les ponts de soufre, qui sont autant de nœuds entre des points éloignés de la chaîne.

Chaque paire d'atomes de soufre appartient à un acide aminé double appelé « cystine ».

Il suffit que le long fil macromoléculaire d'une protéine contienne de la cystine, pour qu'il se trouve comme noué aux points correspondants. Des liaisons de même nature s'établissent aussi entre macromolécules, réalisant une sorte de vulcanisation.

Nombreuses dans la kératine des cheveux, elles leur imposent leur forme droite ou frisée. Pour passer de l'une à l'autre, le coiffeur commence par rompre les liaisons — S — S —, en les réduisant par le sulfure d'ammonium ou par l'acide thioglycolique. Les cheveux roulés finement reçoivent ensuite une application d'eau oxygénée; elle rétablit de nouveaux ponts sulfurés, qui fixent la structure frisée.

Dans le cas de la ribonucléase, on s'est évidemment demandé si son activité est liée à la forme générale de sa macromolécule. En fait, il semble qu'elle soit localisée dans une seule boucle.



Sous ce casque aussi on fait de la chimie macromoléculaire . . .

Doc. Perma

La recherche des séquences d'aminoacides constituant les protéines sera l'une des œuvres maîtresses des années à venir ; elle a déjà été menée à bien pour l'insuline (l'hormone régulatrice de la teneur du sang en glucose, fabriquée par le pancréas), dont la masse moléculaire est inférieure à 6 000, et pour quelques autres protéines plus lourdes, mais sans que leur masse moléculaire dépasse encore 30 000.

La précision avec laquelle ces édifices protéiques sont construits ressort des différences qui ont été mises en évidence entre des ribonucléases apparemment identiques, mais obtenues à partir d'animaux d'espèces différentes.

Des variations de même finesse se révèlent pour des insulines de différentes origines.

LES MÉCANISMES VITAUX, UN A UN, RESSORTISSENT A LA CHIMIE

Le développement d'un être vivant, muni d'organes compliqués, à partir d'un œuf minuscule, aussi bien que sa reproduction fidèle et indéfinie, ont suscité un émerveillement qui date sans doute de l'époque reculée où l'humanité commençait à émerger de l'animalité. Ces processus n'ont cessé de nous apparaître plus clairs à mesure que nos moyens d'observation devenaient plus délicats et notre connaissance de l'infiniment petit plus précise.

aux sources mêmes de la vie : l'A.D.N.

Puisque l'ovule fécondé devient en quelques mois un animal, c'est qu'il contient déjà en puissance tous les éléments de celui-ci. Les progrès de la chimie biologique, conjugués avec les succès des théories de l'hérédité, fondées sur l'hypothèse des gènes, devaient inéluctablement faire rechercher ces éléments dans l'ordonnancement particulier des atomes au sein des cellules de chaque espèce.

Le problème se simplifie un peu, car dans chacune de ces cellules on observe des parties distinctes, aux rôles précis et divers : chromosomes, mitochondries, chloroplastes, ribosomes. Les chloroplastes, porteurs de chlorophylle, se chargent de l'assimilation du gaz carbonique; aux ribosomes échoit la synthèse des protéines. Les chromosomes du noyau président à la reproduction des cellules; ils conservent en dépôt tous les éléments de l'hérédité. Cette fonction, d'après les recherches récentes, revient aux constituants essentiels des chromosomes, les nucléoprotéines, et, dans ces dernières, à la fraction que nous avons dénommée « acides désoxyribonucléiques » (A. D. N.). Un mécanisme qui ne laisse à l'erreur qu'une chance infime en assure la duplication; ils commandent le processus d'évolution de l'œuf, aussi bien que la simple croissance des cellules et les innombrables réactions liées à leur activité.

Pour utiliser une terminologie moderne fort expressive, et dont le lecteur connaît maintenant le sens exact, **les chromosomes contiennent donc toutes les informations nécessaires au déroulement des processus vitaux.** Mais la complication d'une macromolécule peut-elle être suffisamment grande pour qu'elle remplisse les fonctions que nous lui assignons ?

Essayons de nous en rendre compte en construisant l'une d'elles à l'aide de deux motifs élémentaires seulement, représentés symboliquement par A et B. Supposons que chacune n'en contienne que deux; nous aurons quatre possibilités : AA, AB, BA, BB, soit 2^2. Pour construire des molécules à trois motifs, il suffira de placer soit A, soit B à droite des précédentes, ce qui donnera :

AAA ABA BAA BBA
AAB ABB BAB BBB,

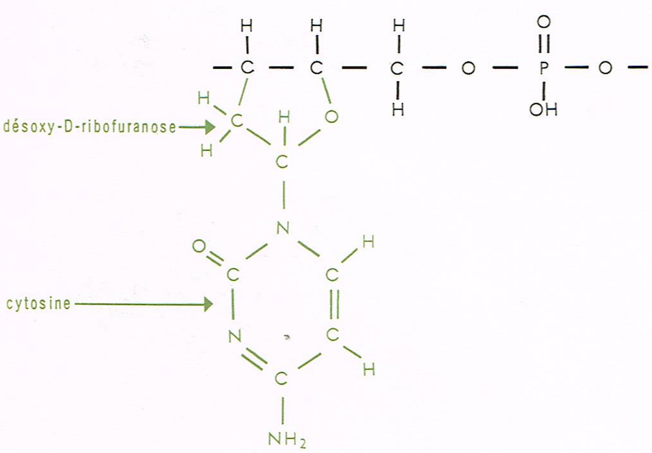
soit $4 \times 2 = 8 = 2^3$ réalisations. On généralisera facilement : une molécule, ou macromolécule, construite avec, au total, n motifs élémentaires de nature A ou B a 2^n possibilités.

Si le nombre des motifs élémentaires s'élevait à trois ou quatre, les combinaisons possibles s'élèveraient à 3^n ou 4^n.

Or, il y a quatre nucléotides susceptibles de former la chaîne des acides désoxyribonucléiques. Pour 10 000 motifs élémentaires, on doit donc prévoir que

$$4^{10\,000} \sim 10^{6\,025} \text{ macromolécules}$$

de ce type sont possibles. Le caractère fabuleux de ce chiffre ressort encore mieux si on l'écrit de façon équivalente : 1, suivi de 6 025 zéros. Il défie l'imagination, et la complexité des phénomènes vitaux la défie aussi; pour l'instant, c'est ce que nous apercevrons de commun entre ces deux ordres de notions. Avant



Un des nucléotides, motifs élémentaires de l'A. D. N. Les autres s'obtiennent en remplaçant la cytosine par les trois cycles azotés : thymine, adénine, guanine.

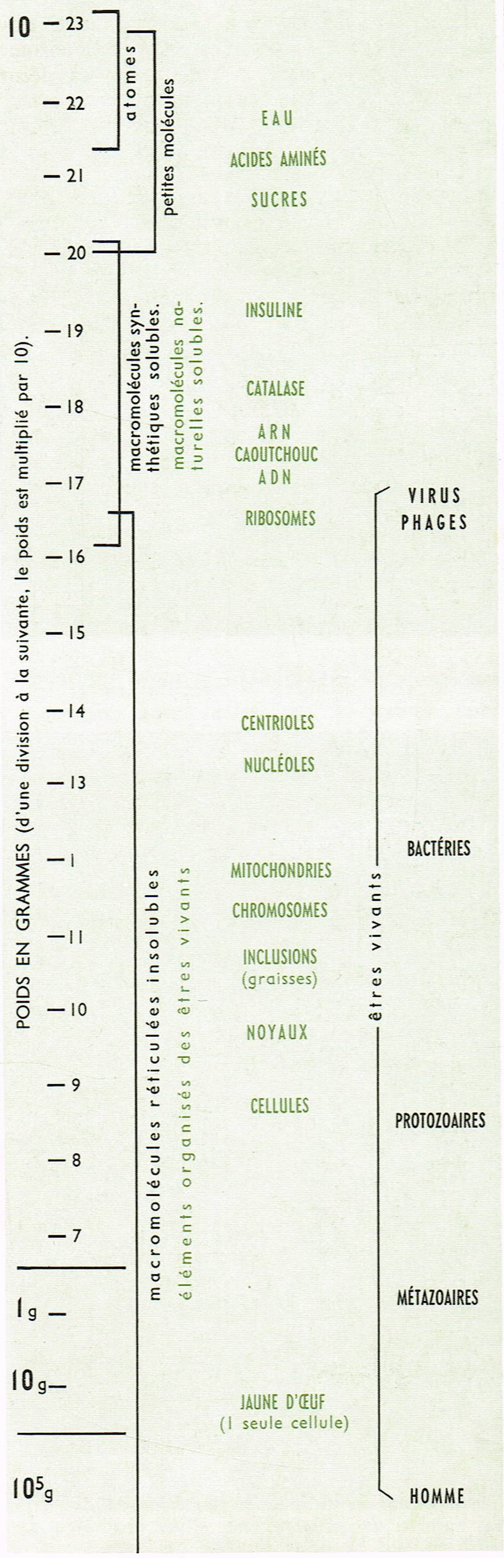


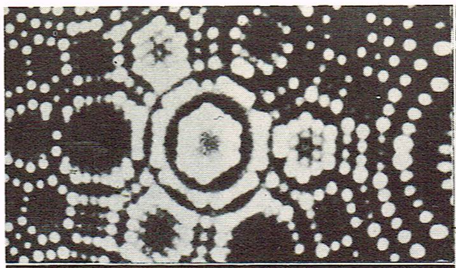
Construction d'un modèle de la molécule d'A. D. N. ***(Une autre représentation, au moyen de boules, en est donnée en couleurs, page 343.)***

Phot. U. S. I. S.

MOLÉCULES ET ÊTRES VIVANTS

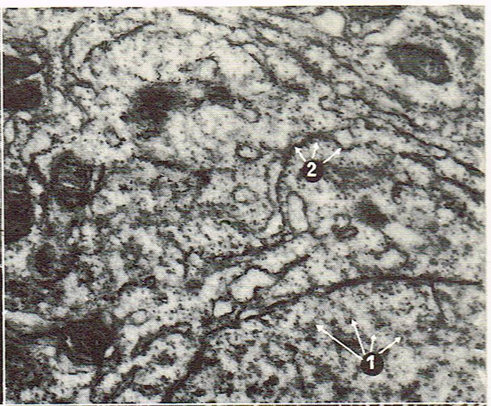
échelle comparative des poids

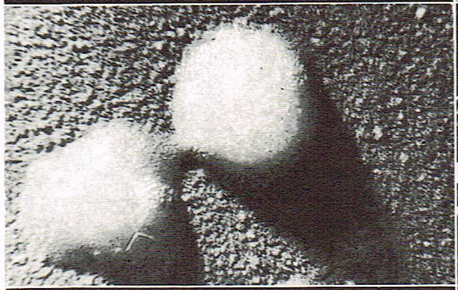




▲ Traces d'atomes de tungstène (gross' 2 000 000).

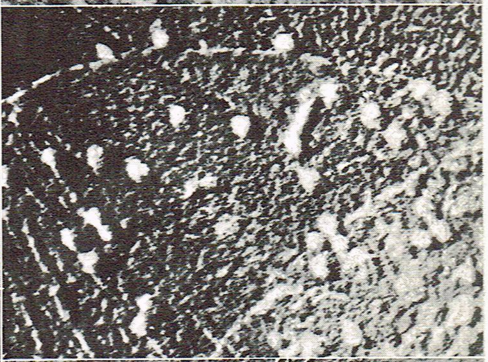
Cellule de foie de poisson : 1. Noyau contenant des grains d'ADN; 2. Grains d'ARN (gross' 22 000). ▶



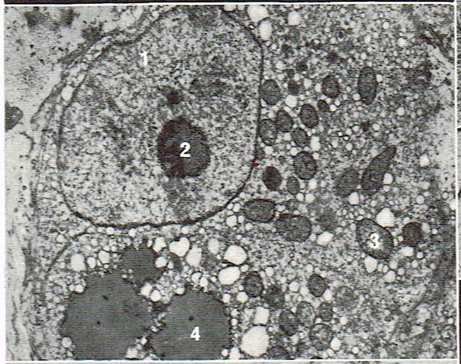


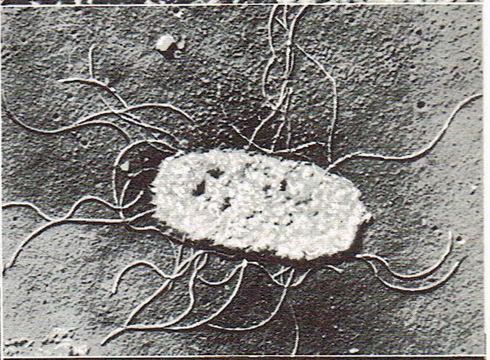
▲ Grains de caoutchouc.

Virus de la grippe. ▶



Cellule de lamproie : 1. Noyau; 2. Nucléole; 3. Mitochondrie; 4. Inclusion (gross' 2 000). ▼

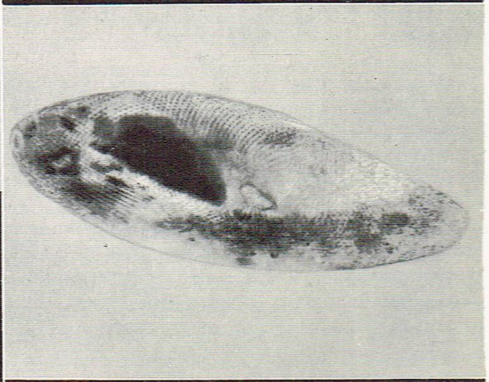




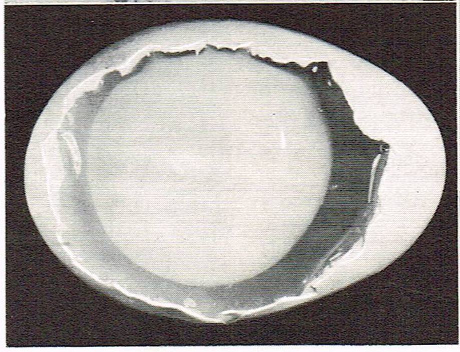
▲ Bactérie (gross' 35 000).

◀ Chromosome isolé d'un protozoaire.





▲ Un protozoaire : *Paramecium caudatum*.



◀ Œuf de poule.

Photos : CCA, CNRS (CMS), Kléber-Colombes, CNRS (P. P. Grassé), Mugard, Noailles.

d'essayer d'apporter quelques précisions sur les rapports qui les lient, dans la mesure où ils se dégagent de l'expérience, nous allons faire plus ample connaissance avec ces acides désoxyribonucléiques, représentés par l'abréviation A. D. N.

Ils ne doivent pas être confondus avec d'autres produits, à vrai dire fort voisins, les acides ribonucléiques, en abrégé A. R. N. En gros, ces derniers se trouvent dans le cytoplasme des cellules, alors que les premiers sont localisés dans le noyau; plus précisément, l'A. D. N., combiné chez les êtres supérieurs à une protéine basique appelée « histone » (1), *entre pour plus de 90 p. 100 dans la constitution des chromosomes, qui ont pu être isolés des noyaux par broyage dans un moulin colloïdal.*

à la découverte de l'A.D.N.

Des échantillons d'A. D. N. d'origines différentes ont été préparés et comparés de multiples fois, mais la plus grande partie des recherches a été effectuée sur le produit extrait du thymus (ris) de veau, matière première abondante et peu coûteuse.

L'ensemble des opérations est effectué à 0 °C, le mieux pour cela étant de s'installer dans une chambre froide. Il faut en effet constamment éviter l'infection microbienne et les phénomènes d'autolyse, c'est-à-dire de destruction spontanée par action de l'eau, catalysée par des enzymes qui sont libérés abondamment dans les tissus, une fois l'animal mort.

Nous avons dit plus haut que l'A. D. N. se trouvait combiné, dans les chromosomes, à une variété de protéine, l'histone. Cette combinaison porte le nom de « nucléoprotéine ». Une première partie des opérations a pour but de l'isoler, en profitant de son insolubilité dans les solutions salées à 1 p. 100 environ.

Le thymus, prélevé dès la mort de l'animal, est congelé dans l'air liquide, puis broyé avec de l'eau salée à 9 g par litre. Après centrifugation et lavages, au moyen de la même solution, il faut séparer l'A. D. N. de l'histone, et aussi d'autres protéines. C'est là le point qui a donné le plus de tracas. Le procédé en faveur met à profit l'action des savons, en l'occurrence de l'un des plus actifs, le dodécylsulfate de sodium. Sa solution dans l'alcool dilué d'eau est agitée avec la nucléoprotéine, puis l'histone précipite par addition d'une quantité importante de sel; ensuite, le sel de sodium de l'A. D. N. se sépare par l'addition d'alcool. Sur cette précieuse substance, nous possédons des renseignements de deux ordres; ils concernent respectivement la nature de ses liaisons chimiques internes et sa morphologie.

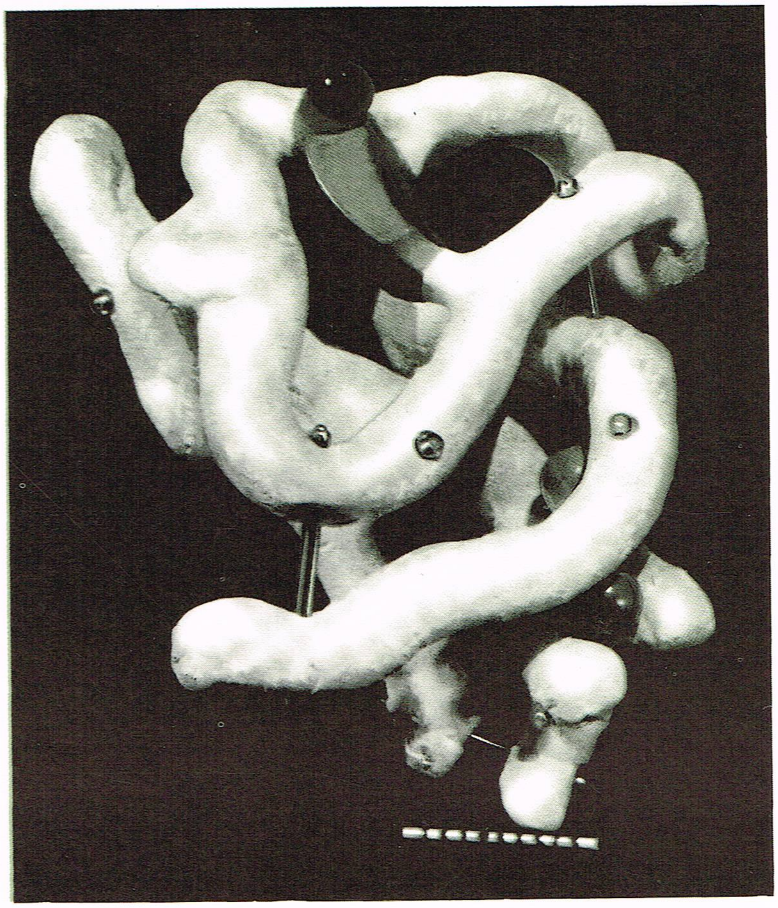
les motifs élémentaires de l'A.D.N. et leur ordre

Pour les chimistes, l'A. D. N. entre dans une catégorie de polyesters phosphoriques dont le représentant le plus simple est celui du propanediol-1-3.

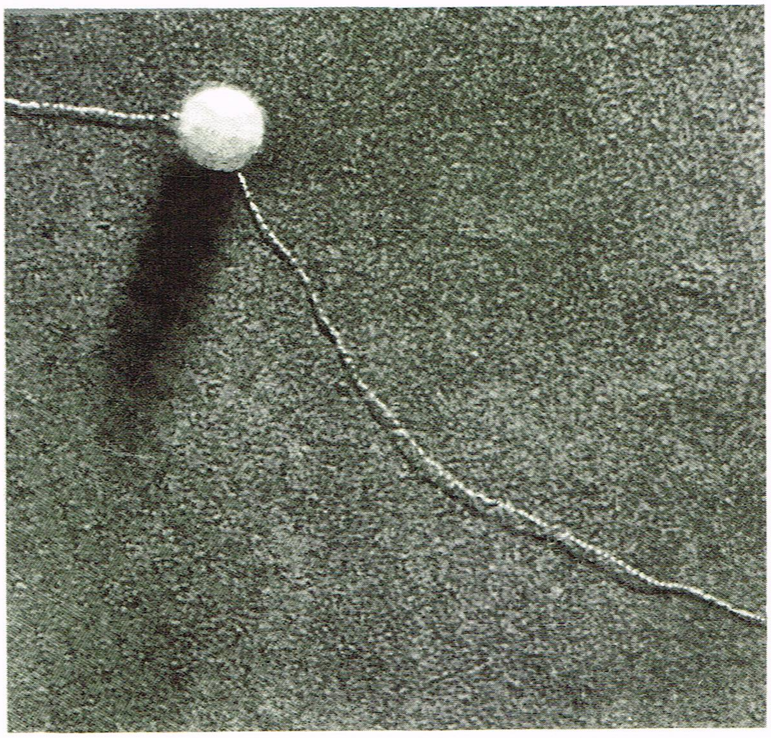
Pour passer de ce polymère à l'A. D. N., il faut d'abord construire sur chaque segment propane le cycle du désoxy-D-ribofuranose, composé appartenant au groupe des sucres et, à ce titre, assez voisin du glucose. Mais ce n'est pas tout, car ce sucre se trouve combiné à un autre composé cyclique, celui-là azoté, choisi parmi quatre espèces différentes : la cytosine (C), la thymine (T), l'adénine (A) et la guanine (G).

---

(1) Le rôle de l'histone n'apparaît pas encore avec certitude; elle fait défaut chez les phages, les virus, les bactéries; l'activité de l'A. D. N. ne lui est donc pas liée nécessairement. Selon des hypothèses récentes, elle servirait à inhiber de façon judicieuse certaines parties de l'A. D. N. des chromosomes. Cela expliquerait pourquoi toutes les cellules d'un même animal se répartissent en groupes différenciés, le foie, le rein, les muscles, la peau, etc., bien que les macromolécules de l'A. D. N. de leurs noyaux soient identiques. C'est que seule la fraction de l'A. D. N. de chacune qui convient à sa fonction n'est pas bloquée par l'histone.

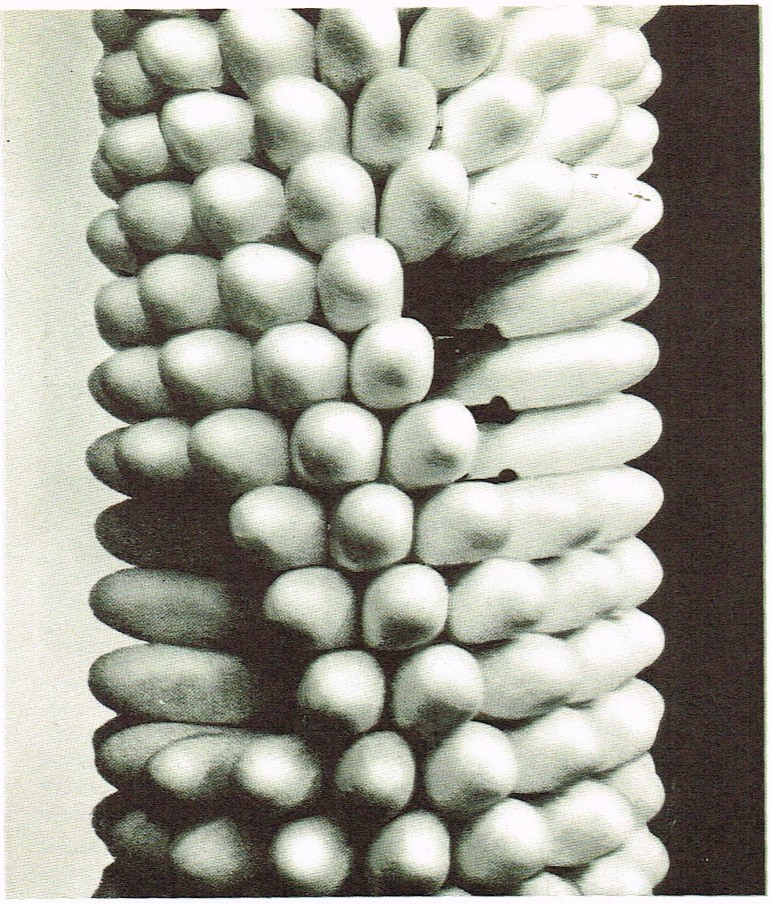


Modèle de macromolécule de myoglobine, d'après Kendrew.

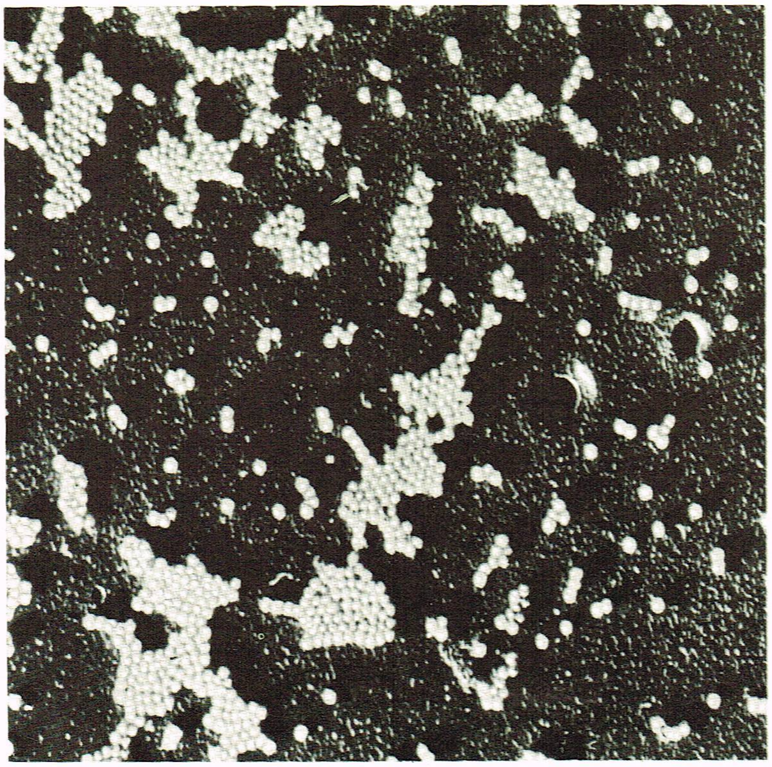


Le filament est une molécule d'A. D. N. provenant du thymus du veau. La sphère de polystyrène, d'un diamètre de 800 angströms, est là pour donner l'échelle.

Microphot. Cecil E. Hall.



Modèle de la structure moléculaire du virus de la mosaïque du tabac, d'après R. Franklin Les protéines (petits corps allongés) sont fixées sur le noyau hélicoïdal d'A. R. N.



Virus de la poliomyélite vu au microscope électronique.

Phot. U. S. I. S.

Finalement, comme nous l'avions annoncé, l'A. D. N. est un polymère construit en enchaînant des motifs élémentaires appartenant à quatre types distincts; chacun d'eux est appelé « nucléotide ». La question de l'ordre de leur succession surgit aussitôt. A propos des protéines, on a vu que la ribonucléase d'une espèce animale résultait de l'enchaînement, dans un ordre parfaitement déterminé expérimentalement, de 124 acides aminés, pris parmi 21 types différents.

Toutes les théories concernant le rôle de l'A. D. N. s'appuient sur l'hypothèse d'un ordre particulier à chaque espèce animale ou végétale dans l'enchaînement des motifs élémentaires.

Nous en avions fait état au début pour montrer que l'A. D. N. possédait une puissance potentielle d'information défiant l'imagination.

Quant à la détermination effective de cet ordre, elle promet d'être ardue; la masse moléculaire de la ribonucléase, dont on connaît bien la structure, est d'environ 13 000; celle de l'hémoglobine, 68 000; tandis qu'une molécule-gramme d'A. D. N. atteint plusieurs millions et comprend un nombre de motifs élémentaires égal à environ 10 000. Mais la tâche que représente son déchiffrement n'est pas pour autant désespérée, car le microscope électronique permettra peut-être de voir et de reconnaître individuellement les motifs élémentaires, après imprégnation par un métal tel que l'or.

la structure spatiale de l'A.D.N.

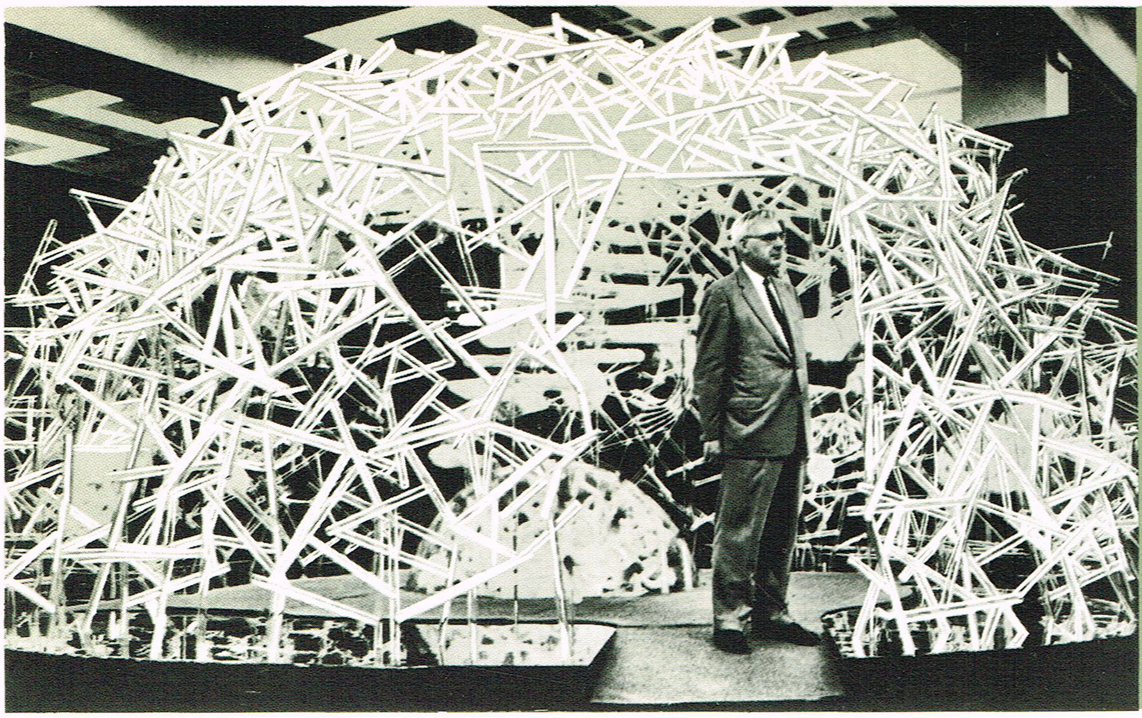
Nous sommes maintenant en possession de tous les éléments permettant de comprendre la structure spatiale de l'A. D. N.

Commençons par enrouler en hélice la chaîne principale, en respectant les angles entre les valences issues d'un même atome; c'est une opération qui a été décrite à propos des polymères isotactiques; elle ne présente pas ici de difficulté plus grande. Il faut maintenant construire une seconde hélice, parallèle à la première; mais — remarque capitale —, ses motifs élémentaires ne sont pas quelconques, car les cycles azotés se placent face à face, l'adénine devant la thymine et la cytosine devant la guanine. Des liaisons hydrogène s'établissent d'un cycle azoté à l'autre, assurant la solidité de l'ensemble, qui forme en quelque sorte une seule macromolécule « en échelle », terme qui fait fortune.

Telle est la macromolécule d'A. D. N., imaginée par Watson et Crick, comme conclusion de leurs célèbres travaux fondés essentiellement sur l'observation des spectres de diffraction de rayons X.

Un détail était sous-entendu dans l'exposé précédent : l'A. D. N. tel que nous l'avons extrait est un sel de sodium des fonctions acides qui subsistent sur les atomes de phosphore. Est-il bien identique à la substance biologiquement active telle qu'elle existe *in vivo?* Sans doute, car il conserve un certain souvenir de cette activité, et des têtes de spermatozoïdes donnent le même diagramme de rayons X.

Une dernière question : quel est le poids de la macromolécule d'A. D. N. et quelle forme affecte-t-elle? Certaines méthodes physiques permettent d'y répondre. Un très grand nombre d'échantillons divers ont été soumis à l'épreuve de la diffusion de la lumière; l'examen des résultats conduit à première vue à une grande confusion, les masses moléculaires moyennes mesurées s'échelonnant de $1,5.10^6$ à 36.10^6; il semble cependant raisonnable d'affirmer l'existence de particules stables de masse moléculaire comprise entre 6.10^6 et 8.10^6, aux erreurs d'expérience près, et cela quelle que soit l'origine du produit : bactéries ou animaux supérieurs. Leur diamètre est voisin de 20 Å; connaissant leur densité, on en déduit que leur longueur atteint 30 000 Å, chiffre paradoxal, puisque les organites cellulaires qui les contiennent mesurent dix fois moins. Il faut donc admettre qu'elles sont repliées sur elles-mêmes, en une douzaine de régions souples de nature inconnue. Cependant les recherches plus récentes



Représentation à trois dimensions d'une cellule humaine (globule rouge du sang) grossie 1 000 000 de fois.

suggèrent que ces longues molécules ne sont que des fragments de celles, mille fois plus longues, qui constituent l'A. D. N. du noyau. Elles ont été brisées du fait des opérations d'extraction et de purification décrites plus haut.

la synthèse de la matière vivante : une énigme

La cellule vivante, qu'elle croisse ou se divise en deux, reproduit sa propre substance. Le mécanisme de cette synthèse est longtemps resté indéchiffrable. Depuis les travaux célèbres de Buchner, qui isola de la levure de bière, par broyage ou macération, des substances capables de provoquer la fermentation du sucre et cependant dénuées de « vie », des enzymes ont été isolés à partir des tissus animaux ou végétaux, en nombre considérable. Ceux qui intéressent les substances macromoléculaires naturelles agissent la plupart dans le sens d'une désagrégation. Il en est ainsi de la pepsine, sécrétée par l'estomac, de la trypsine de l'intestin, de la papaïne végétale, qui hydrolysent les protéines en leurs acides aminés. De même, l'amylase de la salive, du pancréas ou du malt hydrolysent l'amidon en glucose. Longtemps, l'idée a prévalu que ces mêmes enzymes étaient capables, dans certaines conditions, de provoquer les réactions inverses des précédentes, c'est-à-dire des synthèses. Cette conception s'est révélée généralement erronée. Elle était *a priori* incomplète, car elle ignorait le problème essentiel de la spécificité, c'est-à-dire de la reproduction fidèle des macromolécules, avec leurs séquences parfaitement ordonnées de motifs élémentaires.

une première lueur : la synthèse de l'A.D.N.

Or, la seule considération de la structure de l'A. D. N. donne l'impression de saisir l'un des fils directeurs de sa duplication. Ses deux hélices ne sont pas liées très solidement, puisque ce sont des liaisons hydrogène qui les tiennent ensemble.

On conçoit qu'elles puissent se séparer sous l'action de sollicitations de puissance modérée; c'est alors que la synthèse s'effectue. Sur chaque hélice se construit une hélice complémentaire nouvelle, et une nouvelle macromolécule d'A. D. N. en résulte, dont une moitié seulement est réellement neuve. Peut-être — mais ce n'est là qu'une hypothèse — la séparation des deux hélices de la macromolécule mère d'A. D. N. est-elle accompagnée de façon simultanée par les synthèses.

Quelles sont les petites molécules qui viennent ainsi s'agréger en une longue hélice? Ce sont les nucléotides constitutifs de l'A. D. N., combinés à trois molécules d'acide phosphorique. L'intervention de ce dernier est un trait commun à toutes les synthèses que la cellule vivante accomplit avec virtuosité.

Ces quatre triphosphates en solution aqueuse, dans une éprouvette, vont se combiner avec élimination de l'excès d'acide phosphorique, pourvu qu'on leur ajoute un enzyme isolé du bacille *coli*, et, aussi, un peu d'A. D. N. Il y a donc à parier que le nouvel A. D. N. sera très voisin de l'ancien. En fait, l'identité apparaît complète.

les variétés d'acides ribonucléiques, ou A.R.N.

Par leur constitution chimique, les acides ribonucléiques (A. R. N.) s'apparentent étroitement à l'A. D. N. Le désoxy-D-ribofuranose y est remplacé par le D-ribofuranose, qui porte un —OH en plus sur son cycle. D'autre part, au lieu de thymine, on y trouve, combiné au ribose, l'uracile. Les A. R. N. résultent donc eux aussi de la condensation d'un certain nombre de motifs élémentaires appartenant à quatre types distincts, proches parents de ceux de l'A. D. N., appelés également « nucléotides ».

Les matières premières les plus courantes employées pour la préparation des sels de sodium des A. R. N. sont la levure de bière et le foie de bœuf ou de veau. Les procédés ressemblent à celui qui a été décrit pour l'A. D. N. Ces deux catégories de substances se présentent sous des aspects très différents; les premières sont des poudres blanches, les secondes forment des fibres solides. Plus fragiles vis-à-vis des réactifs chimiques, les A. R. N. sont moins bien connus que l'A. D. N.; parmi les données les plus certaines, on peut affirmer que leur composition dépend énormément de leur origine, et qu'il en est de trois sortes; l'une, de poids moléculaire élevé, de l'ordre de 1 à 2 millions, représente 80 p. 100 de l'ensemble (elle entre pour moitié dans la constitution des ribosomes, le reste étant de nature protéique; nous donnerons plus loin quelques indications sur ces particules); les deux autres espèces sont nommées respectivement « A. R. N. messager » et « A. R. N. de transfert ». Ce dernier, appelé aussi « A. R. N. soluble », se distingue par la petitesse relative de sa taille et par sa composition, dans laquelle entrent quelques nucléotides dérivés de cycles azotés nouveaux.

Phot. U. S. I. S.

nouvelles lueurs sur la synthèse de la matière vivante : la synthèse des protéines

D'après ce qui précède, l'A. D. N. est un agent essentiel de sa propre reproduction; il fournit le moule d'une nouvelle macromolécule identique. L'assemblage des acides aminés d'une protéine dans l'ordre convenable est lui aussi assuré par un moule, mais celui-ci n'a rien de commun avec une protéine ; c'est une macromolécule de la variété d'A. R. N. nommée, en raison de sa fonction, « A. R. N. messager ».

L'ensemble du processus se dégage peu à peu, et, à moins de bouleversements toujours possibles dans l'interprétation des données expérimentales, les grandes lignes en sont déjà fixées.

Le lieu de cette synthèse dans la cellule se situe au niveau des ribosomes. Ce sont des granules sphériques d'un diamètre de 100 à 200 Å, révélés par le microscope électronique; ils se trouvent agrégés à la membrane qui, dans chaque cellule, entoure des globules appelés « microsomes ». Les ribosomes contiennent une forte proportion d'A. R. N., lié à une protéine, environ 50 p. 100; dans une seule cellule d'*Escherichia coli* leur nombre est de 10 000. Afin de les séparer, on soumet les cellules, d'abord désintégrées par broyage, irradiation aux ultrasons, etc., à des centrifugations à vitesses croissantes. Lorsque la force centrifuge atteint 500 à 1 000 fois la pesanteur (500 à 1 000 g), les noyaux se sédimentent. Pour 10 000 à 15 000 g, c'est le tour des mitochondries et de fragments divers. Enfin les ribosomes se déposent pour 20 000 à 100 000 g.

En l'absence de toute cellule vivante, les aminoacides se condensent en protéines s'ils se trouvent mélangés aux ribosomes, additionnés d'acide adénosine-triphosphorique (A. T. P.) et aussi de liquide cytoplasmique. Les observations ultérieures ont montré la richesse de celui-ci. Il contient, entre autres, des enzymes et la variété d'A. R. N. dite « messager ».

chaque molécule d'acide aminé est prise en charge par l'A.R.N. de transfert qui convient à son espèce

A chaque variété d'acide aminé, et nous avons vu qu'elles sont au nombre d'une vingtaine, correspondent un A. R. N. de transfert et un enzyme spécifique.

Une molécule d'acide aminé, la phénylalanine par exemple (en abrégé Phe), est chargée sur la macromolécule d'A. R. N. adéquate, à laquelle elle se combine sous l'influence de l'enzyme spécifique. Admirons ici comment les risques d'erreur, inséparables de tout mécanisme, se trouvent réduits par un moyen qui n'est pas sans analogie avec le geste familier du caissier qui ouvre un coffre-fort : il faut à la fois composer le chiffre correct et utiliser la bonne clef. A cette réaction participe l'acide adénosine-triphosphorique (A. T. P.).

Chaque molécule d'acide aminé, guidée par son A. R. N. de transfert, rejoint sa place sur l'A. R. N. messager et s'intègre dans la protéine.

Continuons à suivre la molécule de phénylalanine (Phe), maintenant fixée sur son A. R. N. de transfert particulier. Celui-ci la transporte à la place exacte qui lui revient sur une immense matrice; là elle se lie aux deux acides aminés voisins, rangés aussi avec la même précision inéluctable, tandis que les A. R. N. de transfert, dont le rôle est terminé, se détachent et s'éloignent, prêts pour un nouveau transport.

Quelle est la nature de cette matrice à laquelle nous venons de faire allusion?

On a évidemment soupçonné les ribosomes de jouer ce rôle, puisque la condensation a lieu à leur contact, mais c'était une erreur. Une telle fonction revient à une troisième espèce d'A. R. N., d'ailleurs éphémère, appelée « A. R. N. messager ».

Chaque variété de cet A. R. N. messager est fabriquée, suivant les besoins du moment, pour ainsi dire « à la demande », par un organisme central qui n'est autre que l'A. D. N. des chromosomes, et selon un processus dont nous parlerons plus loin.

le déchiffrement du code génétique à peu près achevé

On pressent qu'une relation étroite existe entre l'ordre des nucléotides dans une macromolécule d'A. R. N. messager et l'ordre des acides aminés qu'il impose à sa protéine fille.

Elle évoque les dictionnaires codes ou grilles qui permettent de passer d'une langue à une autre, ou d'une écriture conventionnelle à une autre. La tâche offerte ici aux chercheurs est donc le déchiffrement d'un code, qui mérite bien le nom de « code génétique ».

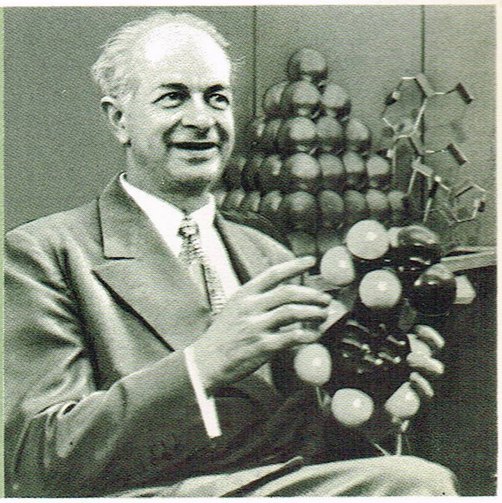
Ce déchiffrement a pu commencer, car nous avons acquis le pouvoir extraordinaire de préparer synthétiquement des A. R. N. messagers de structure simple et bien établie.

Sous l'influence d'un enzyme appelé « polynucléotide phosphorylase », les mononucléotides, en effet, s'assemblent. Ainsi, en partant de l'acide uridylique (U) dérivé de l'uracile, on obtient un A. R. N. artificiel, l'acide polyuridylique, de structure particulièrement simple, puisqu'il est constitué par une succession de motifs identiques. Introduit dans le milieu complexe décrit précédemment, il provoque la formation d'un polypeptide dérivé uniquement de la phénylalanine, une condition nécessaire étant évidemment que, parmi les acides aminés présents, figure celui-là. On en déduit qu'une séquence — — — UUUUU — — — dans l'A. R. N. de transfert se traduit par « phénylalanine ».

Expérimentons maintenant un A. R. N. artificiel obtenu à partir d'acide uridylique (U) auquel on a mélangé une petite proportion d'acide adénylique (A), nucléotide dérivé de l'adénine.

La protéine isolée contient, à côté de la phénylalanine, des traces d'isoleucine. A une séquence — — — UUUAUUU — — —, dont il faudra préciser le nombre des termes, correspond donc l'isoleucine.

Pour une proportion plus grande d'acide adénylique, le polynucléotide formé doit contenir, en plus, des paires AA noyées dans les U, c'est-à-dire des séquences



Linus Carl Pauling, physicien américain, prix Nobel de chimie (1954) et de la paix (1963), l'un des meilleurs spécialistes dans le domaine des macromolécules organiques.

Phot. U. S. I. S.

——— UUUUAAUUU ———.

Or, on voit corrélativement apparaître dans le polypeptide un troisième acide aminé, l'asparagine.

Les expériences de ce genre ont été multipliées; elles ont permis d'attribuer à chacun des 21 acides aminés naturels la séquence de nucléotides qui lui correspond dans l'A. R. N. messager. Il apparaît que ces séquences caractéristiques, dont quelques-unes, parmi les plus certaines, sont indiquées par le tableau ci-dessous, comportent trois motifs seulement.

phénylalanine	UUU	tyrosine	UUA
sérine	UUC	valine	UUG
proline	UCC	leucine	UUG
isoleucine	UUA	tryptophane	UGG

Plusieurs « triplets » formés des mêmes nucléotides peuvent différer, dans leur signification, par l'ordre de ceux-ci, qui demeure inconnu. Par exemple, il est impossible de distinguer, jusqu'à présent, entre UUA, UAU et AUU.

C'est la raison pour laquelle on a dû se résigner à écrire UUA aussi bien pour l'isoleucine que pour la tyrosine.

Le nombre des séquences différentes de 3 motifs, pris parmi 4 variétés, est de $4^3 = 64$.

Comme on connaît seulement 21 acides α-aminés naturels, il faut envisager la possibilité d'un code « dégénéré », plusieurs séquences correspondant à un seul acide aminé, et certaines n'ayant aucun sens. Si les séquences ne comportaient que 2 motifs, leur nombre $4^2 = 16$ serait insuffisant; pour 4 motifs, il serait exagérément grand : $4^4 = 256$.

On précisera encore que la macromolécule d'A. R. N. messager se « lit » en la subdivisant en triplets à partir d'un point fixe, ces triplets ne chevauchant pas l'un sur l'autre. Les macromolécules, relativement petites, d'A. R. N. de transfert, chacune portant son acide aminé particulier, reconnaissent donc leur place respective grâce à un groupe de trois nucléotides, complémentaires probables d'un triplet de l'A. R. N. messager.

Simultanément, sans confusion d'ordre, les aminoacides ainsi transportés se combinent. Alors interviennent d'autres enzymes, ainsi que les ribosomes, selon un processus non encore élucidé.

Ce sont les A. R. N. artificiels qui ont permis les beaux travaux de déchiffrement du code de la synthèse des protéines, mais l'emploi d'A. R. N. de transfert naturels a donné lieu aussi à de magnifiques recherches; mêlés aux éléments d'une cellule étrangère, ils provoquent la reproduction des protéines qui leur sont propres. Nous comprenons mieux ainsi comment agissent les virus et les phages.

un gros livre héréditaire

Passons à l'organisme responsable qui délègue des messagers; l'A. D. N. du noyau, on l'a deviné, en est une pièce essentielle. C'est lui qui constitue le gros livre héréditaire sur lequel est inscrit le programme de la vie de la cellule. Un bactériophage ou un virus, avec une chaîne d'A. D. N. comportant 50 000 motifs, contient environ 15 000 mots, soit 30 pages d'un livre moyen. Les chromosomes d'une cellule somatique humaine représentent l'équivalent de 48 volumes de 12 000 pages chacun.

La plupart du temps, cette énorme encyclopédie est fermée, mais, lorsque le besoin s'en fait sentir, quelques pages s'ouvrent et sont reproduites sous la forme de macromolécules d'A. R. N. messager. Le mécanisme de cette copie ressemble probablement à celui de la duplication de l'A. D. N., c'est-à-dire que l'hélice servant de matrice engendre son propre négatif. Les messagers migrent vers les ribosomes, provoquent à leur niveau la synthèse des protéines attendues et sont détruits.

Pourquoi certaines pages s'ouvrent-elles et au terme de quel processus? De la solution de ce problème, les grandes lignes commencent à être perçues.

Supposons qu'une cellule vivante, une bactérie par exemple, se trouve plongée dans un milieu riche en sucre. La fabrication des enzymes protéiques nécessaires à la fermentation va commencer. Chacun possède un « modèle », constitué par un segment particulier de l'A. D. N. contenu dans le chromosome. Chacun des segments s'identifie à un gène, dit « gène de structure », et tous ceux qui concourent à l'accomplissement d'une fonction, ici la fermentation du glucose, se trouvent sous la dépendance d'un autre segment du même filament d'A. D. N., appelé « opérateur ». A l'ensemble de l'opérateur et des gènes de structure qu'il commande on donne le nom d' « opéron ». En l'absence de glucose, l'opérateur est combiné à un « répresseur » dont les travaux récents montrent qu'il s'agit d'une protéine spécifique. Lorsque le glucose apparaît, il se combine au répresseur, le soustrayant à l'opérateur, qui, ainsi libéré, déclenche alors l'activité de l'opéron.

Nous voilà arrivés à la pointe extrême des efforts tendant à réduire la vie à des mouvements d'atomes. Certains phénomènes vitaux essentiels apparaissent maintenant comme des réactions chimiques, dont les fins détails se préciseront peu à peu. N'oublions pas, cependant, que des problèmes considérables subsistent. Si nous entrevoyons le mécanisme de la réplication de l'A. D. N., aucune donnée expérimentale sérieuse ne nous apprend comment sont apparues sur notre planète les premières de ces macromolécules. En d'autres termes, quelle est l'origine de la vie? Rechercher les réactions chimiques qui, à partir de substances minérales telles que l'azote, l'eau, l'anhydride carbonique, pourraient conduire aux ancêtres des êtres unicellulaires que nous connaissons, c'est faire un pari hasardeux, mais rien ne permet d'affirmer qu'il ne sera pas gagné un jour.

La mémoire pose une autre énigme; on peut concevoir qu'elle ait pour support des modifications dans la structure des macromolécules appartenant aux cellules nerveuses. Pour l'instant, nous ne possédons aucun renseignement sur la nature de ces substances enregistreuses, non plus que sur les modifications correspondant à l'inscription d'un souvenir, mais ne désespérons pas d'acquérir quelque jour des lumières sur cette question.

La vie, avec toutes ses manifestations conscientes, deviendra-t-elle donc, en dernier ressort, un chapitre de la chimie macromoléculaire? Nous ne croyons pas à une telle possibilité, à moins que des modifications profondes ne soient apportées aux lois de la chimie, à l'échelle des macromolécules. C'est alors que l'étude de ces dernières aura réalisé ses plus précieuses conquêtes.

Les très basses TEMPÉRATURES

par

M. VEYSSIE-COUNILLON et G. KUHN

Auprès des très hautes températures, réalisées ou réalisables, et qui se chiffrent par millions de degrés, le domaine du froid peut paraître, à première vue, bien limité; il n'est pas question, en effet, d'aller au-delà du zéro absolu, qui se situe à quelque 300 °C seulement au-dessous des températures qui nous sont habituelles. En fait, le zéro absolu est une limite inaccessible; on ne peut que s'en approcher de plus en plus, mettant en jeu des techniques spécifiques toujours plus complexes. On peut dire que les très basses températures constituent véritablement, à l'heure actuelle, une branche de la physique : la cryogénie. Et, après en avoir délimité le champ d'action, après avoir décrit les clés de ce royaume, nous verrons ce que la science du froid peut apporter à l'homme, tant sur le plan de la connaissance que dans le cadre de réalisations industrielles.

LES GRANDES DATES DE LA CRYOGÉNIE

1792 - Liquéfaction de l'ammoniac (— 30 °C) par von Marum.

1825 - Liquéfaction de nombreux gaz par compression (Faraday).

1853 - Découverte de l'effet Joule-Thomson.

1869 - Liquéfaction du gaz carbonique (— 78 °C) par Andrews.

1877 - Première liquéfaction de l'oxygène (— 183 °C) par Cailletet et Pictet.

1883 - Liquéfaction de l'oxygène (— 183 °C) et de l'azote (— 196 °C) par von Wroblewski et Olszewski. Premières traces d'hydrogène liquide.

1898 - Dewar liquéfie l'hydrogène (— 252 °C).

1902 - Claude réalise la liquéfaction par détente avec travail extérieur.

1908 - Liquéfaction de l'hélium (— 269 °C, 4,2 °K) par Kammerlingh Onnes. En abaissant la pression sur le liquide, il atteint près de 1 °K.

1911 - Kammerlingh Onnes découvre les supraconducteurs.

1926 - Debye et Giauque proposent la désaimantation adiabatique électronique pour atteindre des températures très inférieures à 1 °K. Giauque obtient 0,25 °K en 1933; cette technique permet maintenant d'atteindre 10^{-3} °K.

1956 - Utilisation de la désaimantation adiabatique nucléaire pour descendre en dessous de 10^{-3} °K.

1961 - London et Mendoza suggèrent l'idée d'un réfrigérateur par dilution continue de l'isotope hélium 3 dans l'hélium 4.

1966 - Réalisation expérimentale de cet appareil à dilution dans plusieurs laboratoires; les meilleures performances se situent déjà vers 10^{-2} °K.

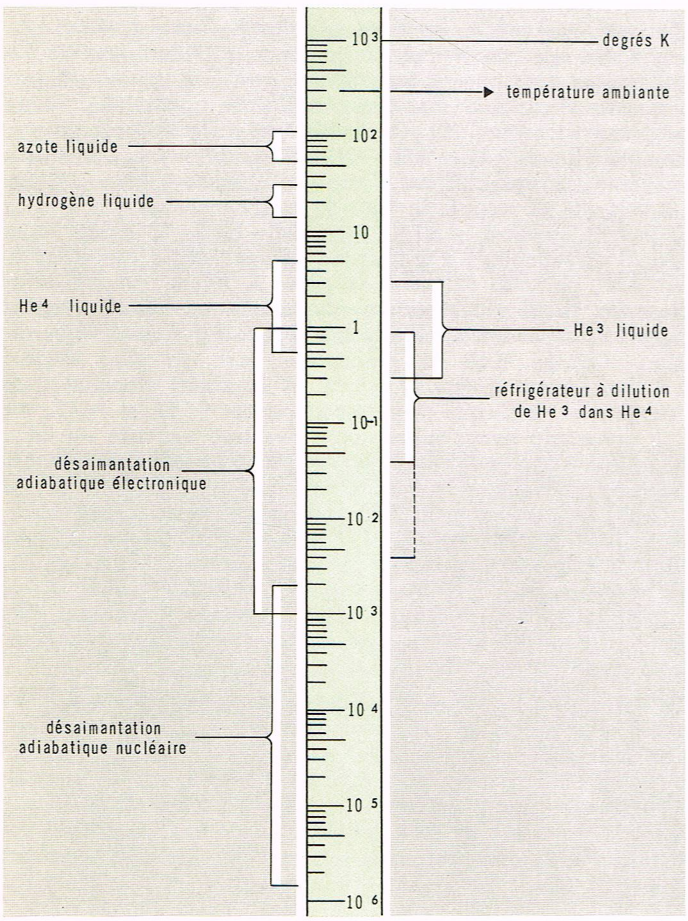
La limite inférieure, nous l'avons dit, en est parfaitement déterminée, c'est le zéro absolu, à — 273°15 dans l'échelle Celsius. On peut, par contre, en fixer assez arbitrairement une borne supérieure. Au sens propre, on pourrait parler de cryogénie dès qu'apparaît pour nous la sensation de froid. En pratique, il faut distinguer deux domaines assez différents : celui des basses températures, entre 0 °C et la température d'ébullition de l'azote, sous pression normale, soit — 196 °C, et celui des très basses températures, qui s'étend, au-delà, jusqu'au zéro absolu. On voit que les records naturels de froid enregistrés à la surface de la terre, environ — 80 °C, font piètre figure. Il devient alors naturel d'abandonner la graduation Celsius (définie, rappelons-le, par 0 °C, point de fusion de la glace, et 100 °C, point d'ébullition de l'eau, sous pression normale) pour l'échelle Kelvin (°K). La valeur absolue du degré Kelvin est égale à celle du degré Celsius, mais la graduation part du zéro absolu; 0 °C correspond alors à 273,15 °K, et les points d'ébullition de l'oxygène et de l'azote sont respectivement 77 °K et 90 °K. Et l'on comprend mieux alors quelle aventure nouvelle est cette exploration du froid. On a pu obtenir, récemment, des températures de l'ordre de 10^{-6} °K, c'est-à-dire quelques centaines de millions de fois plus petites que la température ambiante. Exprimés ainsi en termes de rapport, les records des cryogénistes sont aussi spectaculaires que ceux des spécialistes des hautes températures, et surtout ils nous font augurer qu'il s'agit là d'un monde radicalement différent, où ne sauraient se conserver les propriétés de la matière et les lois qui nous sont, aux températures usuelles, familières.

Ce que nous avons appelé domaine des basses températures est maintenant courant dans l'industrie et même dans la vie quotidienne : réfrigérateurs, congélateurs, transports frigorifiques, etc. Nous parlerons donc surtout du domaine des très basses températures; bien qu'exploré depuis plus de cent ans (voir le tableau des grandes dates de la cryogénie), il est encore peu connu et reste l'apanage des laboratoires de recherches et de quelques industries spécialisées. Ses débouchés, bien que connus ou soupçonnés très tôt, à partir des découvertes fondamentales, n'ont cependant trouvé que récemment leurs applications : fusées propulsives à l'hydrogène liquide, bobines supraconductrices, etc. Mais on peut maintenant s'attendre à la généralisation rapide de ces techniques.

comment descendre l'échelle des températures

La première grande date de l'histoire de la cryogénie est la découverte faite simultanément par J.-B. Joule et W. G. Thomson, plus tard lord Kelvin, et qui porte leurs noms conjugués. L'effet Joule-Thomson fut vraiment la clé des basses températures. Si on laisse un gaz préalablement comprimé se détendre à travers un petit orifice, il subit un effet thermique. On avait généralement observé une élévation de température, mais si, dans les conditions initiales, la température du gaz est assez basse, et sa pression suffisamment élevée, c'est un refroidissement qui a lieu pendant cette détente. Et c'est ainsi qu'un grand nombre de gaz, qualifiés jusqu'alors de « permanents », ont pu être liquéfiés. Il s'agit en pratique d'une opération, la « cascade »; une certaine quantité de gaz se refroidit par détente, permet d'abaisser la température initiale d'une autre fraction, qui, après détente, se refroidit davantage, et l'opération se poursuit ainsi jusqu'à la liquéfaction. Mais, condition primordiale, il faut dès le départ une température assez basse pour que l'effet thermique se produise dans la bonne direction. C'est ainsi qu'on doit pré-refroidir l'hydrogène avec de l'azote liquide, et pré-refroidir l'hélium à l'aide de l'hydrogène. Si ce procédé est, aujourd'hui encore, largement utilisé, d'autres méthodes ont été mises au point. Ainsi la détente avec travail extérieur, due à G. Claude : le gaz, convenablement isolé du milieu extérieur, se refroidit parce qu'on l'oblige à fournir du travail pendant qu'il se détend.

Nous trouvons un peu plus loin, récapitulées dans un tableau, les caractéristiques des principaux liquides cryogéniques; ce sont des fluides à bas points d'ébullition, avec, tout au bas de l'échelle, l'hélium, qui bout à 4,2 °K, sous pression normale. On peut d'ailleurs, en pompant sur ces liquides et en diminuant la tension de vapeur, obtenir des températures plus basses encore, la limite étant, sauf pour l'hélium, la température de solidification. Le problème de l'hélium est particulier : c'est seulement sous pression élevée qu'on peut l'obtenir à l'état solide. Quand, au contraire, on diminue la pression au-dessus d'un bain d'hélium, la température décroît, bien sûr, mais, à 2,2 °K, il se produit un phénomène brutal : l'hélium subit une transformation de nature et devient superfluide ; c'est le « point lambda ». L'hélium superfluide a un comportement très différent d'un liquide normal; par exemple, sa conductivité thermique devient presque infinie. Par malheur, l'hélium a tendance à « grimper » le long des parois des récipients, et il y a alors une évaporation considérable. Il faut utiliser des pompes à très fort débit. En prenant, en outre, certaines précautions (réduction du col des vases, double bain d'hélium), on peut arriver à 0,7 °K.



Les clés des très basses températures

On peut d'ailleurs éviter ces inconvénients, et « gagner » quelques dixièmes de degré, en utilisant un isotope rare de l'hélium, l'hélium 3. D'une part, son point d'ébullition sous pression normale est inférieur à 3,3 °K, mais, surtout, il ne subit pas de transformation superfluide et on peut obtenir, par pompage, une température de 0,3 °K. Mais, cette fois-ci, nous n'avons plus aucun autre fluide de remplacement. Alors, que fait-on? A ce stade-là, c'est dans la matière à l'état solide que l'on trouve de nouveaux agents frigogènes : on utilise, en effet, des cristaux paramagnétiques. Ce sont des composés où figurent des ions porteurs de moments magnétiques (famille du fer, famille des terres rares). Un tel sel paramagnétique est tout d'abord refroidi aussi bas que possible à l'aide d'hélium liquide. On lui applique alors un champ magnétique, qui oriente les moments magnétiques des ions, et qui porte ainsi le système, toujours en contact avec le bain d'hélium, dans un état ordonné. Si, alors, on isole thermiquement le sel et si on annule le champ magnétique, c'est-à-dire si on procède à une désaimantation adiabatique, l'échantillon se refroidit; en effet, l'entropie qui représente l'état d'ordre du système reste constante dans une telle opération adiabatique. Il faut donc que le système se refroidisse pour compenser, par une diminution de température, la disparition du facteur d'ordre qu'était le champ magnétique.

On peut ainsi obtenir des températures de l'ordre du millième de degré! On peut même prétendre descendre à 10^{-6} °K, en appliquant cet effet à des moments nucléaires (du cuivre, par exemple) pré-refroidis à 10^{-3} °K. Mais la désaimantation nucléaire reste encore à l'état de prouesse.

On se tourne, actuellement, vers une nouvelle méthode, qui met à profit la chaleur d'évaporation en régime continu de l'hélium 3 dans l'hélium 4, et permet déjà d'atteindre $2,10^{-2}$ °K. Cette méthode possède un triple avantage : elle est beaucoup moins onéreuse; elle permet de faire des mesures à température constante, alors que, dans une méthode de désaimantation adiabatique, on assiste à un réchauffement continu de l'échantillon à partir de la température minimale atteinte; enfin, l'absence de toute source de champ magnétique évite les champs qui, même faibles, peuvent perturber les mesures physiques.

les thermomètres cryogéniques

Comment mesurer des températures aussi basses? Le système de référence est le thermomètre à gaz, fondé sur la loi de Mariotte, qui relie la pression, le volume et la température d'un gaz parfait. En pratique, on doit heureusement n'avoir qu'assez rarement recours à ces dispositifs délicats; ils servent essentiellement à étalonner des thermomètres secondaires, d'emploi plus commode, et qui sont de natures assez variées, puisque, en principe, toutes les grandeurs physiques qui varient de façon reproductible avec la température peuvent être utilisées. Nous ne citerons que les plus répandues.

Quand on travaille dans un bain de fluide cryogénique, il est facile d'en mesurer la tension de vapeur et d'en déduire la température à l'aide de tables établies et admises internationalement. Pour l'hélium, par exemple, il existe de telles tables qui définissent la température à 10^{-4} °K près.

Au-dessous de 1 °K (ou de 0,3 °K pour l'hélium 3), il ne peut plus être question de tension de vapeur de bain liquide. On mesure alors la susceptibilité d'un sel paramagnétique; la théorie montre qu'elle varie en raison inverse de la température; il suffit donc d'extrapoler la loi de variation établie dans une région où l'on dispose de mesures thermométriques précises.

On peut également suivre la résistance électrique de certains corps, méthode très agréable pour obtenir la température au niveau d'un échantillon. Les résistances thermométriques métalliques ont une sensibilité médiocre au-dessous de 20 °K. Par contre, les semi-conducteurs présentent, à très basse température, une variation très rapide de la résistivité : une résistance de 50 Ω à la température ambiante peut passer, par exemple, à 100 Ω vers 20 °K, à 1 000 Ω vers 4 °K, et à 100 000 Ω vers 1 °K. On se sert, actuellement, soit de simples résistances radio (graphites), soit d'éléments au germanium.

Les thermocouples, enfin, d'un emploi commode, peuvent être utilisés, mais leur sensibilité décroît très vite avec la température. L'un des plus sensibles dans la région qui nous intéresse est le système cuivre-or-cobalt.

la cryogénie et la physique fondamentale

La cryogénie est devenue un outil indispensable au progrès de la physique moderne. Rappelons tout d'abord que c'est l'accès aux très basses températures qui a permis la découverte de deux phénomènes aussi particuliers qu'importants : la superfluidité de l'hélium, qui apparaît en dessous de 2,2 °K et que nous avons déjà évoquée, et la supraconductivité. Ce sont là des problèmes dominants de la science fondamentale qui n'ont cessé de mériter l'attention d'innombrables chercheurs tant expérimentaux que théoriciens et suscitent encore à l'heure actuelle de fructueuses controverses. Mais on peut dire, de façon plus large, que les très basses températures constituent le domaine de prédilection du physicien du solide. Elles sont en effet indispensables pour mettre en évidence dans la matière les phénomènes quantiques qui ne se manifestent que si l'énergie thermique est réduite à un degré identique ou inférieur à celui de l'énergie quantique résiduelle. Nous n'en donnerons qu'un exemple. A température ambiante, la chaleur spécifique des éléments, pris à l'état solide, obéit à la loi de Dulong et Petit, c'est-à-dire qu'elle équivaut, pour un atome-gramme de substance, à une constante, soit 6 calories/degré$^{-1}$. Cette loi s'interprète très bien dans le cadre classique de la répartition de l'énergie entre les 6 N degrés de liberté du système (N, nombre d'Avogadro). Mais si la température décroît, on constate expérimentalement que la chaleur spécifique s'écarte de cette valeur constante pour s'annuler en même temps que la température, ce qu'Einstein a interprété, de façon qualitative tout au moins, sur la base de la quantification de l'énergie de vibration des atomes.

Nous ne pouvons donner ici une nomenclature complète des phénomènes de physique qui n'apparaissent qu'aux basses températures; disons simplement qu'à l'heure actuelle il n'y a pas un seul laboratoire au monde où l'on puisse vraiment étudier la physique de la matière condensée sans avoir un équipement cryogénique.

de la science appliquée à l'industrie

L'*azote liquide* est devenu actuellement un des liquides de base de nombreuses industries. Il sert de liquide de pré-refroidissement lorsqu'on veut travailler à plus basse température, mais la grosse consommation d'azote liquide est destinée à des applications industrielles. On peut citer des domaines extrêmement divers : traitements thermiques en métallurgie (trempe); assemblage de deux pièces par contraction de l'une d'elles refroidie à l'azote liquide; applications spatiales (chambres de simulation, par exemple); installations de cryopompage, pièges à azote liquide, fournitures d'azote pour les applications chimiques. Il faut dire là que la liquéfaction de l'air est un procédé pour séparer l'oxygène de l'azote. On alimente alors, à partir de telles unités, des installations de fabrication de nombreux produits azotés, tel l'ammoniac. Applications biologiques : conservation de cellules, de tissus et même d'organes, stockage et transport de spermes pour l'insémination artificielle, traitements de maladies du cerveau ou de l'œil; transports de produits alimentaires, etc.

L'oxygène. — La production mondiale d'oxygène liquide peut être évaluée actuellement à 100 000 tonnes par jour au moins.

CARACTÉRISTIQUES DES FLUIDES CRYOGÉNIQUES

					ÉTAT SOLIDE		ÉTAT LIQUIDE			ÉTAT GAZEUX	
NOM	Symbole	Poids moléculaire (g)	Température critique (°K)	Pression critique (bar)	Température du point triple (°K)	Chaleur latente de solidification au point triple (cal/mole)	Poids spécifique sous 1,013 03 bar (kg/dm³)	Température ébullition (°K sous 1,013 03 bar)	1 litre liquide = X, Nm³ de gaz à 0 °C et 1,013 03 bar	Chaleur latente de vaporisation sous 1,013 03 bar (cal/mole)	Chaleur nécessaire pour vaporiser 1 litre de liquide (cal)
Hydrogène normal	H_2	2,016	33,26	12,97	13,95	28,0	0,070 79	20,39	0,789	216,0	7 585
Hydrogène para	H_2	2,016	32,98	12,94	13,84	28,03	0,070 77	20,27	0,787	214,8	7 540
Hélium	He	4,003	5,25	2,29			0,125 2	4,22	0,746	19,8	619
Néon	Ne	20,183	44,45	27,2	24,55	80,1	1,204	27,25	1,328	410,0	24 458
Azote	N_2	28,016	126,2	33,93	63,15	172,0	0,808 1	77,35	0,646	1 336,0	38 536
Oxygène	O_2	32,000	154,75	50,4	54,32	105,6	1,147	90,15	0,803	1 632,0	58 497
Argon	A	39,944	150,69	48,62	83,80	280,0	1,392	87,29	0,78	1 553,8	54 159



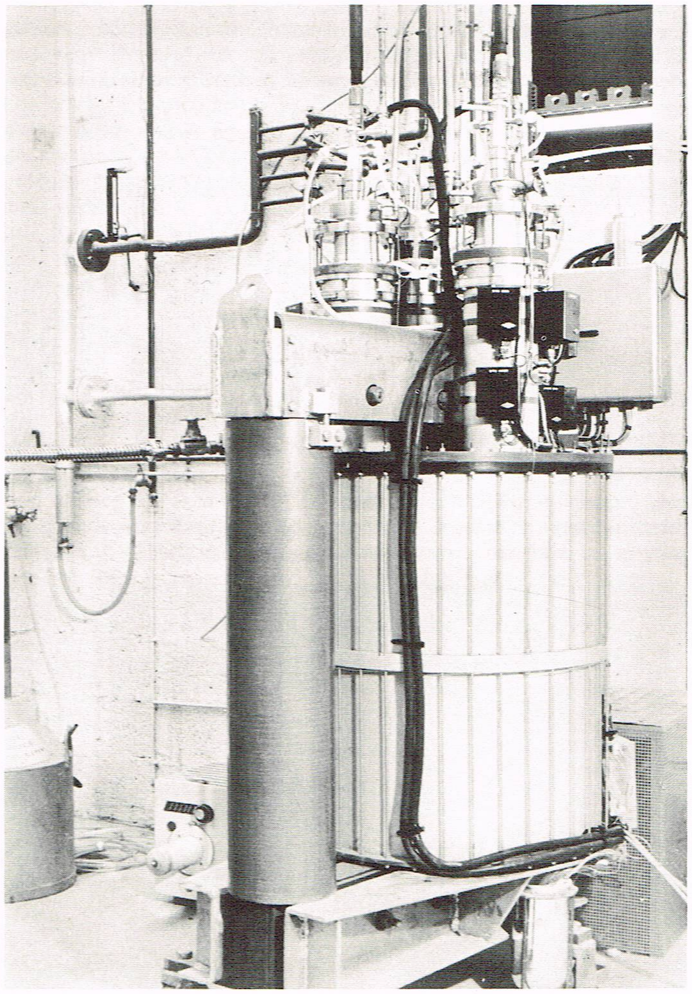
Traitement de la maladie de Parkinson et de l'épilepsie par cryocautérisation.

Si, dans ce cas aussi, la liquéfaction sert surtout à séparer l'oxygène de l'azote et à fournir l'oxygène pour les innombrables applications de ce gaz à l'état pur (soudure oxyacétylénique, métallurgie, chimie, médecine, etc.), son usage à l'état liquide se développe rapidement. En Lorraine, l'oxygène insufflé dans la fonte pour préparer l'acier par le procédé Bessemer est ainsi transporté à l'état liquide par pipe-lines depuis l'usine de liquéfaction jusqu'aux aciéries. Pour les fusées, l'oxygène liquide est un comburant presque idéal.

L'*hélium* a été découvert, pour la première fois, dans les raies du Soleil, d'où son nom. L'atmosphère en contient trop peu pour qu'on puisse en obtenir de grandes quantités par liquéfaction de l'air. La production d'hélium est actuellement faite à partir de certains gaz naturels de pétrole qui en contiennent une fraction appréciable. Les sources pour le monde occidental sont essentiellement aux États-Unis. Comme pour les autres gaz, les principales applications de l'hélium ne sont pas directement dans le domaine des basses températures, et la liquéfaction n'est ici qu'un moyen d'isoler l'hélium, qui est ensuite utilisé à l'état gazeux dans de nombreux domaines, par exemple pour gonfler des ballons d'observation atmosphérique, créer des atmosphères inertes en métallurgie, etc. De nombreuses applications résultent du faible poids atomique de l'hélium. C'est ainsi qu'en remplissant les pneus d'un avion avec de l'hélium au lieu de l'air, on arrive à économiser le poids d'un passager. Mais, pour nous, c'est le précieux fluide de base de la cryogénie, précieux à tous les sens du mot : si un litre d'azote liquide revient environ à 1 F le litre, il faut compter en Europe de l'ordre de 30 F par litre d'hélium liquide, et le gaz est soigneusement récupéré. Ce n'est encore rien si on pense que l'isotope He_3 produit par réaction nucléaire à partir de lithium revient, lui, à 1 000 F le litre de gaz, sous la pression normale.

L'*hydrogène* est également un liquide cryogénique fondamental. Il est récupéré dans de nombreux gaz naturels ou est obtenu à partir de méthane, par exemple. Les applications de l'hydrogène sont extrêmement nombreuses, et la production atteint plusieurs centaines de tonnes par jour. Les très gros consommateurs sont les fusées propulsées à l'hydrogène liquide et les chambres à bulles, réalisations qui sont décrites par ailleurs dans ce livre.

L'hydrogène est un fluide bien moins cher que l'hélium liquide, et, à ce titre, il est très employé malgré les risques d'explosion qu'il peut présenter. Il a aussi une chaleur de vaporisation bien supérieure à celle de l'hélium (voir le tableau des caractéristiques des fluides cryogéniques). Une application très récente, mais peut-être de grand avenir, est l'utilisation en électrotechnique et en électricité des hyperconducteurs.



Maquette d'un cryotransformateur, réalisée conjointement par les Sociétés Alsthom, Pechiney et Air liquide. Caractéristiques de fonctionnement : conducteur en aluminium très pur refroidi à 20 °K, puissance 690 kVA, circuit magnétique à température ambiante, cryostat isolant en verre-résine epoxy, tension 1 080 × 2 160 V. On économise les deux tiers de conducteur et le volume total est la moitié de celui d'un transformateur classique.

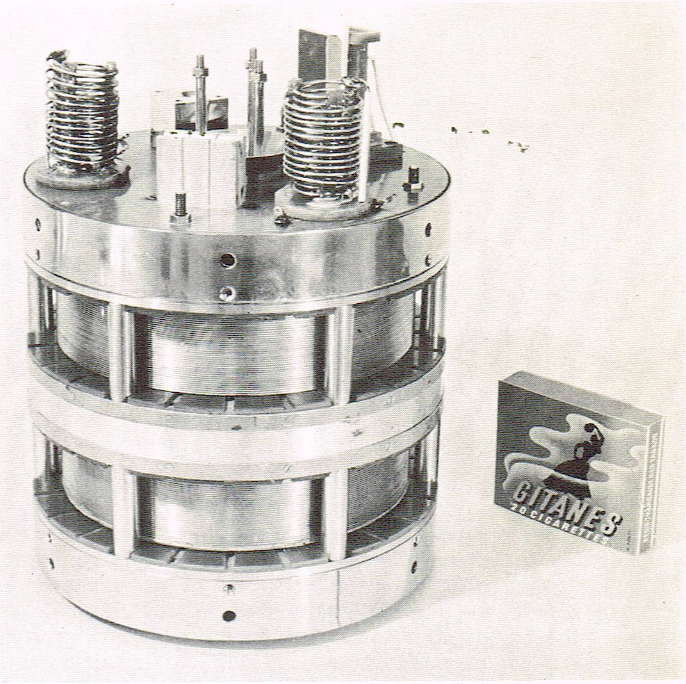
Phot. Air liquide, Lacheroy.

hyperconducteurs et supraconducteurs

Les métaux bons conducteurs, tels que le cuivre, l'aluminium, etc., voient leur résistance électrique très nettement diminuée lorsqu'on les refroidit à 20 °K, s'ils sont suffisamment purs. On gagne ainsi un facteur très important, d'une part dans les pertes et, d'autre part, dans le volume des matières nécessaires. Ces conducteurs purs à basses températures, qu'on a baptisés hyperconducteurs, font déjà l'objet de nombreuses recherches dans le domaine des transports du courant (lignes hyperconductrices), celui des transformateurs, ou même des machines tournantes. Pour de très gros transformateurs, on pense ainsi gagner un facteur 2 sur la taille totale du transformateur, en économisant la moitié de fer et les deux tiers de conducteur. L'installation cryogénique est évidemment coûteuse, et, au moins dans l'état actuel, on ne peut envisager ces améliorations que pour de très grosses unités.

De nombreux métaux, lorsqu'ils sont refroidis à quelques degrés Kelvin, ont la remarquable propriété de perdre toute résistance électrique. Cet effet, découvert dès 1911 par Kammerlingh-Onnes, n'a trouvé d'applications que très récemment. En effet, une des premières idées qui avaient été envisagées pour utiliser les supraconducteurs était de faire des bobines pour créer des champs magnétiques. L'ennui est qu'il existe un champ magnétique au-dessus duquel la supraconductibilité disparaît. Depuis une dizaine d'années, de nombreuses études ont permis de fabriquer des alliages dont les champs critiques atteignent des dizaines de kOe et même pour l'alliage Nb_3Sn : 180 000 Oe, alors que les métaux purs n'ont que des champs critiques de l'ordre de quelques centaines ou milliers d'Oe. Ces bobines fonctionnent dans l'hélium liquide, mais la puissance nécessaire pour produire des champs intenses est nulle. Les bobines supraconductrices connaissent actuellement un développement considérable. On est en train d'équiper avec des bobines supraconductrices même de très grosses installations, par exemple les aimants de chambres à bulles ou les aimants de déflexion des accélérateurs. L'installation cryogénique et le prix élevé des matériaux supraconducteurs sont largement compensés par la puissance de fonctionnement nulle.

Une autre application : les mémoires supraconductrices pour les machines à calculer. Les temps de commutation de ces mémoires sont nettement inférieurs à ceux des mémoires magnétiques classiques, et il est possible que, dans un proche avenir, de nombreux calculateurs soient équipés de tels ensembles à mémoires supraconductrices. L'élément de base de ces mémoires est le cryotron, ensemble de deux filaments supraconducteurs croisés l'un sur l'autre; le champ magnétique créé par le courant circulant dans l'un pouvant détruire le courant circulant dans l'autre en le rendant normal. La mémoire est parfaite du fait de l'inexistence de résistivité électrique, donc de la persistance infinie d'un signal envoyé dans la mémoire.



Bobines de Helmholtz supraconductrices en câble d'alliage niobium-zirconium, produisant 40 kilogauss et utilisées pour des expériences de diffraction de neutrons.

Bien que moins spectaculaires, de nombreuses autres applications se développent rapidement (gyroscope supraconducteur, bolomètres, etc.).

En conclusion, nous dirons que la grande aventure du froid, commencée il y a plus de cent ans, est loin d'être terminée. Si les températures de l'azote liquide et même de l'hydrogène sont bien commercialisées, le domaine de l'hélium est également en train d'enfoncer les portes de l'industrie, avec, notamment, les applications des supraconducteurs. Les laboratoires de recherche pure ne seront cependant pas à court d'ouvrage, car en dessous de 1 °K on peut dire que presque tout reste à faire, et on peut prévoir que les contributions apportées à la science et à l'industrie seront encore innombrables.

Phot. Air liquide

OPTIQUE NON LINÉAIRE

par

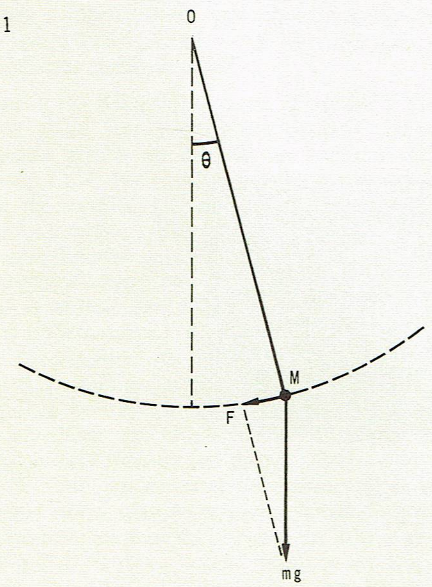
Maurice FRANÇON

Dans beaucoup d'expériences de physique, la relation qui existe entre deux quantités n'est linéaire qu'en première approximation. En d'autres termes, la réponse (sortie) peut ne pas être proportionnelle à l'excitant (entrée). Par exemple, lorsqu'on écarte un pendule OM d'un angle θ de sa position d'équilibre, ce pendule est soumis à une force F qui est, en première approximation, proportionnelle à l'angle θ. Si m est la masse du pendule, g l'accélération de la pesanteur, mg est le poids du pendule; celui-ci est soumis à une force $F = mg\theta$. La relation entre F et θ est linéaire, mais cela n'est plus vrai si l'angle θ n'est pas petit.

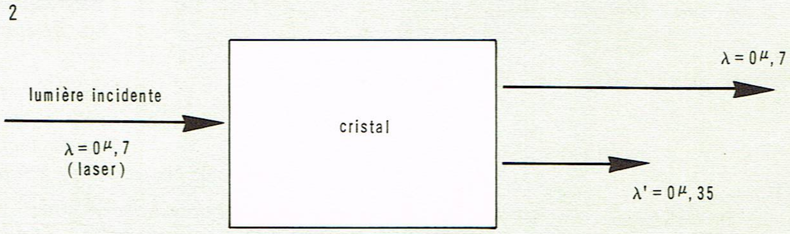
Des phénomènes de non-linéarité peuvent aussi s'observer en optique, par exemple entre la polarisation d'un cristal et le champ électrique auquel est soumis ce cristal. Lorsqu'un cristal est éclairé par une source ordinaire, le champ électrique n'est jamais très grand et la polarisation du cristal est proportionnelle au champ électrique. Cela n'est plus vrai si le champ devient plus intense; c'est justement ce qui se produit avec des lasers capables de donner des champs d'une très grande intensité. Avec un laser à rubis, on peut obtenir des champs de 10^5 V/cm dans le faisceau émergent, et en concentrant celui-ci sur une surface de 1 micron de diamètre, le champ électrique peut atteindre 10^8 V/cm, qui est l'ordre de grandeur des champs atomiques présents dans la structure atomique d'un cristal. Dans ces conditions, la polarisation du cristal n'est plus proportionnelle au champ électrique. Ce phénomène se traduit de la façon suivante : en concentrant le faisceau d'un laser à rubis sur un cristal, à la sortie de celui-ci on retrouve naturellement la radiation incidente, mais cette radiation est accompagnée d'une nouvelle radiation de longueur d'onde deux fois plus petite. Par exemple, si le laser émet une longueur d'onde de 0,7 micron, à la sortie du cristal on observe une radiation de même longueur d'onde, très intense, plus une nouvelle radiation peu intense de longueur d'onde 0,35 micron. Cette nouvelle radiation est appelée harmonique n° 2, l'harmonique n° 1 étant la radiation incidente.

Ce phénomène fondamental est appelé phénomène d'optique non linéaire. La proportion d'énergie de la radiation incidente qui est transformée pour produire l'harmonique n° 2 dépend du cristal choisi et de l'intensité de la radiation incidente.

1



2



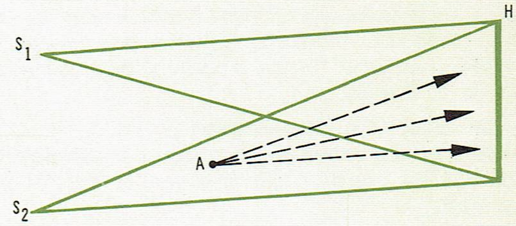
1. Lorsqu'on écarte un pendule OM d'un angle θ de sa position d'équilibre, ce pendule est soumis à une force F proportionnelle à l'angle θ en première approximation. La relation entre F et θ est linéaire.

2. Un faisceau laser incident de longueur d'onde $\lambda = 0{,}7$ donne naissance, à la sortie du cristal, à un faisceau de même longueur d'onde, et à un deuxième faisceau, beaucoup moins intense, de longueur d'onde deux fois plus petite $\lambda = 0{,}35$. Ce phénomène est appelé phénomène d'optique non linéaire.

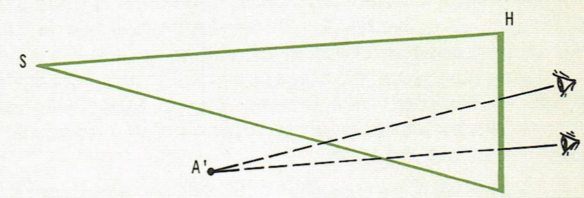
Application à l'holographie

L'hologramme est une photographie sur laquelle on ne voit absolument rien qui puisse révéler l'objet photographié. En éclairant convenablement l'hologramme et en le regardant par transparence, on observe une image de l'objet reconstitué en trois dimensions. La technique des hologrammes a pu être mise au point grâce à l'invention des lasers. La figure 1 montre comment s'effectue la prise de l'hologramme. On dispose de deux sources, S_1 et S_2, parfaitement cohérentes. Elles proviennent d'une source unique (un laser). Nous décrirons plus loin le montage qui permet de produire les deux sources ponctuelles S_1 et S_2 avec un seul laser. Une plaque photographique H est située à une certaine distance de S_1 et S_2 qui peut être égale à 1 m, par exemple. Pour simplifier, on considère comme objet un point A, tel qu'une petite bille de verre. La bille A est éclairée par l'un des deux faisceaux, le faisceau provenant de S_2, et diffracte de la lumière qui provient de S_1. L'ensemble des faisceaux est recueilli par la plaque H. On a un véritable interféromètre : la lumière de l'un des deux faisceaux éclaire l'objet A (faisceau issu de S_2) et l'autre faisceau (faisceau issu de S_1) va directement sur la plaque H.

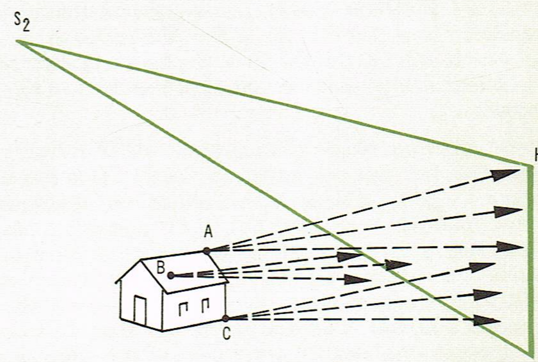
Développons la plaque ainsi exposée et observons à travers la plaque, comme l'indique la figure 2. L'hologramme H est éclairé par transparence avec une source ponctuelle S (provenant d'un laser) dans les mêmes conditions que précédemment. A travers la plaque H l'observateur voit une image virtuelle A′ de A dans l'espace situé précisément là où se trouvait la bille. Remplaçons le point A par un grand nombre de points A, B, C, etc. : l'ensemble peut constituer un objet à trois dimensions. Si l'objet est opaque, il est éclairé comme le montre la figure 3. Le phénomène est le même pour tous les points. Un objet à trois dimensions est donc reconstitué dans l'espace (*fig.* 4). Ce n'est pas un relief stéréoscopique. Les deux yeux observent l'image comme un objet ordinaire. Les schémas des figures 2 et 4 sont en réalité incomplets : en plus de l'image virtuelle, il y a une seconde image, réelle celle-là, et qui se trouverait à droite de l'hologramme sur les figures 2 et 4.



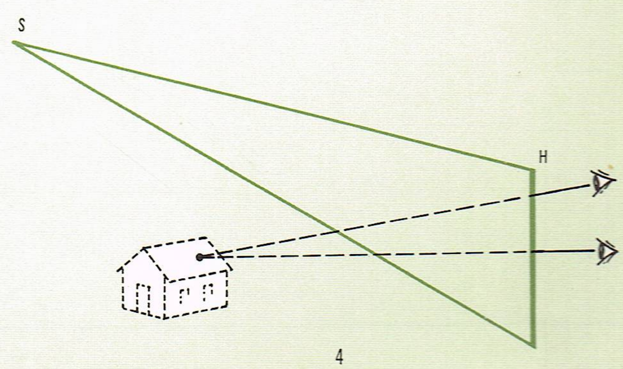
1



2



3



4

1. Enregistrement d'un hologramme. L'objet A est un simple point qui diffracte un peu de lumière provenant de la source S_2. Les deux sources S_1 et S_2 sont cohérentes (elles proviennent, en réalité, d'une même source : un laser).

2. Après développement, l'hologramme est éclairé dans les mêmes conditions que lors de l'enregistrement : la source S est la même que la source S_1 de la figure 1.

3. Dans le cas général d'un objet quelconque, par exemple une petite maison, l'objet est éclairé par une source non représentée sur la figure. Cette source est cohérente avec la source S_2 qui éclaire directement la plaque H. Les différents points A, B, C,... diffusent de la lumière sur H, qui enregistre les interférences des vibrations envoyées par S_2 et les différents points de l'objet.

4. Après développement, on éclaire l'hologramme H par une source S identique à la source S_2 de la figure 3 et qui occupe la même position par rapport à l'hologramme H. A travers l'hologramme, l'observateur voit une image virtuelle qui reconstitue l'objet en trois dimensions.

La figure 5 montre un exemple de montage permettant d'enregistrer un hologramme. La source L provient d'un laser : elle éclaire l'objet diffusant A et, grâce au miroir plan réfléchissant M, une partie Σ de la lumière va directement sur la plaque photographique H. Après développement et si on éclaire la plaque par un faisceau identique au faisceau Σ utilisé à l'enregistrement, on obtient une image virtuelle A_1 et une image réelle A'_1 (*fig.* 6). A travers la plaque H, l'observateur voit une image virtuelle A_1 qui reconstitue l'objet en trois dimensions. L'image réelle A'_1 peut être photographiée directement.

Notons que l'hologramme recueille les informations sur toute sa surface, et, si on le casse, on retrouve pratiquement les mêmes phénomènes avec un seul morceau. C'est un peu comme si on casse une lentille : on peut former l'image d'un objet avec un seul morceau de la lentille.

Lorsque les différences des distances des points A, B, C, etc., à la plaque H sont assez grandes (de 10 à 50 cm), la théorie montre qu'il faut une lumière très monochromatique, et c'est pourquoi on utilise un laser.

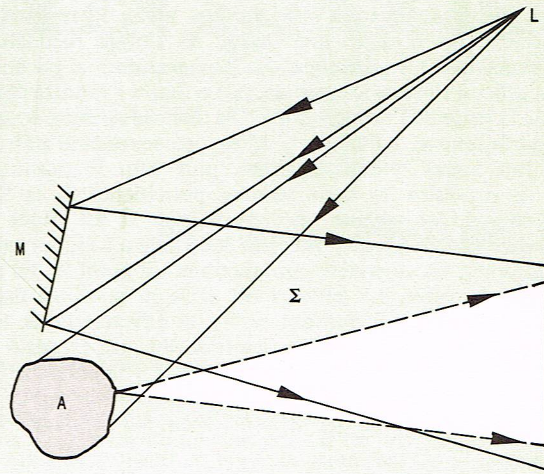
Grâce à certaines propriétés dues à l'épaisseur de l'émulsion photographique, on peut réaliser des hologrammes en couleur. Il faut bien noter que les plaques utilisées sont encore des plaques noir et blanc et que c'est un phénomène d'interférence qui permet de reproduire les couleurs. Pour enregistrer un hologramme en couleur, on éclaire l'objet avec trois longueurs d'onde convenables (rouge, vert, bleu). Après développement de la plaque dans les conditions habituelles, si on éclaire en lumière blanche, seules les trois longueurs d'onde qui ont été employées à l'enregistrement reconstituent des images. On a donc trois images qui se superposent exactement et reproduisent l'objet en trois dimensions avec ses couleurs.

Les applications de l'holographie sont très nombreuses. Citons le problème important de la reconnaissance des formes : il s'agit de reconnaître dans un texte des lettres déterminées et leurs positions. La solution est obtenue grâce à un dispositif dont l'élément fondamental est un hologramme. On peut donc envisager la machine à lire comme une extrapolation. Bien entendu, la reconnaissance de la position et de la forme d'un objet ne se limite pas aux lettres d'un texte. On peut appliquer la méthode à n'importe quel type d'objet dans lequel on recherche une information déterminée.

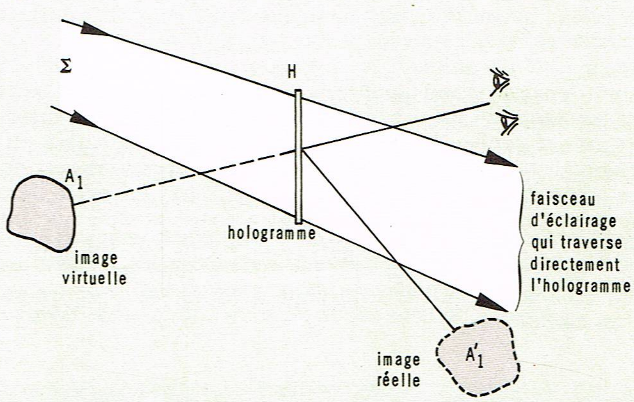
Une particularité remarquable de l'holographie est la suivante : si l'enregistrement est fait avec une longueur d'onde λ et la reconstitution avec une longueur d'onde λ', on obtient un grandissement qui est multiplié par le rapport λ'/λ. Il est possible d'envisager un microscope à rayons X et la reconstitution avec une longueur d'onde du spectre visible. Un pouvoir séparateur analogue à celui du microscope électronique pourrait être obtenu, mais le microscope à rayons X est encore du domaine du futur. L'holographie apporte aussi des solutions élégantes à des phénomènes d'interférométrie et on peut envisager des développements importants dans beaucoup d'autres domaines de la science et de la technique.

5. La source L provient d'un laser. Elle éclaire directement la plaque photographique H par l'intermédiaire d'un miroir plan M. Elle éclaire aussi l'objet A qui diffuse de la lumière sur H. Le faisceau Σ qui éclaire directement H et les faisceaux diffusés par les différents points de l'objet A sont bien cohérents puisqu'ils proviennent de la même source lumineuse ponctuelle L.

5



6



6. A la reconstitution, l'hologramme est éclairé par un faisceau Σ identique à celui qui est utilisé lors de l'enregistrement. L'hologramme reconstitue deux images, l'une virtuelle A_1, qui peut être observée directement, l'autre réelle A'_1, qui peut être photographiée.

— édition 1970 —

Imprimerie LAROUSSE, 1 à 9, rue d'Arcueil, Montrouge (Hauts-de-Seine).
Juillet 1964. — Dépôt légal 1964-3e. — N° 4 501. — N° de série Editeur 4 723. — IMPRIMÉ EN FRANCE (*Printed in France*). — 16 110 B-10-69.

ENCYCLOPÉDIE FRANÇAISE

Fondateur : Anatole de Monzie

TOME I **L'OUTILLAGE MENTAL**
pensée, langage, mathématique
directeurs : A. Rey, A. Meillet, P. Montel.

TOME II **LA PHYSIQUE**
directeur : L. de Broglie.

TOME III **LE CIEL ET LA TERRE**
directeurs : A. Danjon, P. Pruvost, J. Blache.

TOME IV **LA VIE** EN 2 VOLUMES
fondement, maintien, reproduction
directeur : P.-P. Grassé.

TOME V **LES ÊTRES VIVANTS**
directeurs : P. Lemoine, R. Jeannel, P. Allorge.

TOME VI **L'ÊTRE HUMAIN**
directeur : R. Leriche.

TOME VII **L'ESPÈCE HUMAINE**
directeur : P. Rivet.

TOME VIII **LA VIE MENTALE**
directeur : H. Wallon.

TOME IX **L'UNIVERS ÉCONOMIQUE ET SOCIAL**
directeur F. Perroux.

TOME X **L'ÉTAT**
directeurs : E. Faure, L. Trotabas.

TOME XI **LA VIE INTERNATIONALE**
divisions et unité du monde actuel
directeur : J. de Bourbon Busset.

TOME XII **CHIMIE, science et industries**
directeur : A. Kirrmann.

TOME XIII **INDUSTRIE - AGRICULTURE**
directeurs : J. Capelle, P. Chouard.

TOME XIV **LA CIVILISATION QUOTIDIENNE**
directeur : P. Breton.

TOME XV **ÉDUCATION ET INSTRUCTION**
directeur : C. Bouglé.

TOMES XVI ET XVII **ARTS ET LITTÉRATURES**
directeur : P. Abraham.

TOME XVIII **LA CIVILISATION ÉCRITE**
directeur : J. Cain.

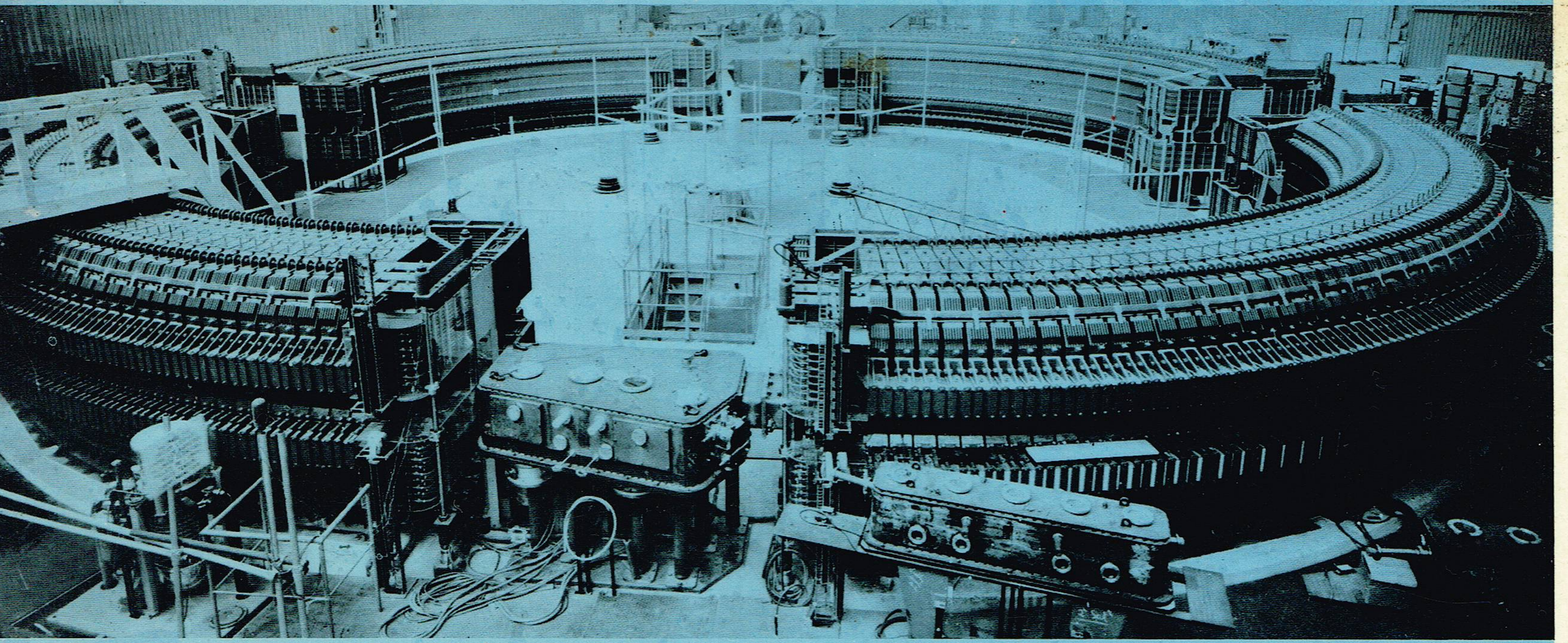
TOME XIX **PHILOSOPHIE - RELIGION**
directeur : G. Berger.

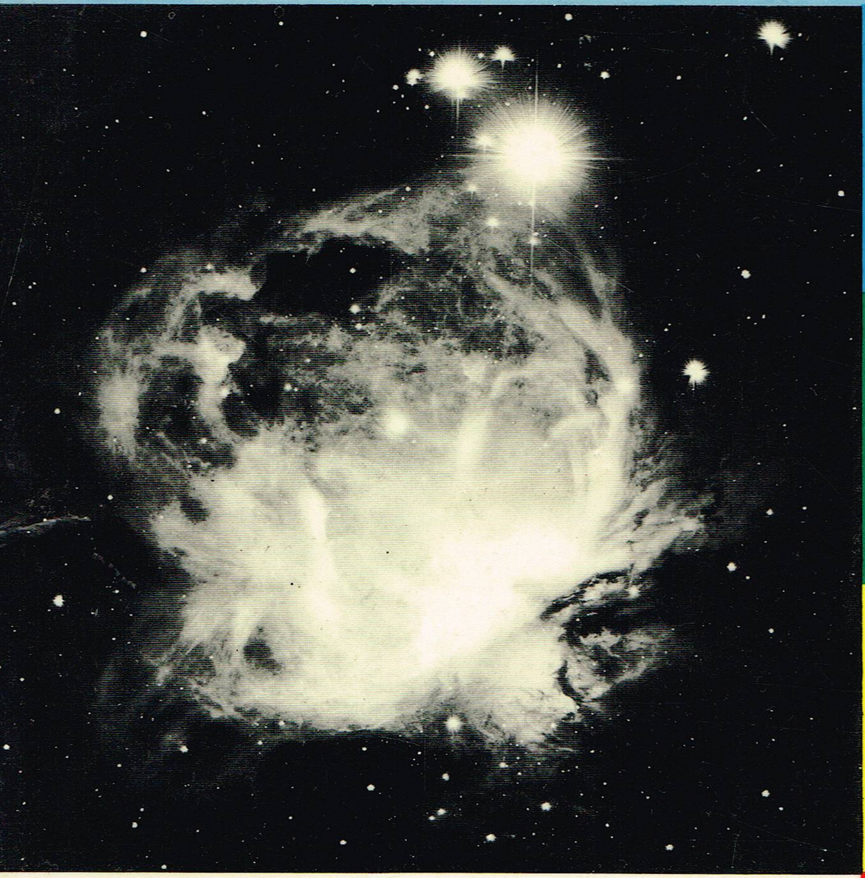
TOME XX **LE MONDE EN DEVENIR**
histoire, évolution, prospective
directeurs : G. Berger, P. Renouvin.

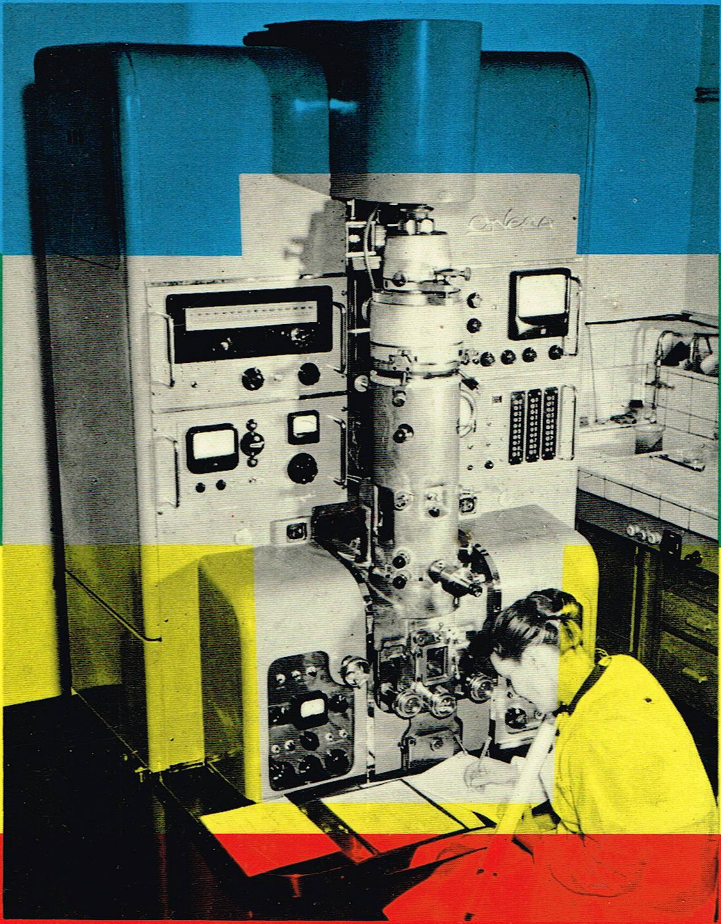
Il existe des "Cahiers d'Actualité et de Synthèse" complétant certains des volumes ci-dessus.

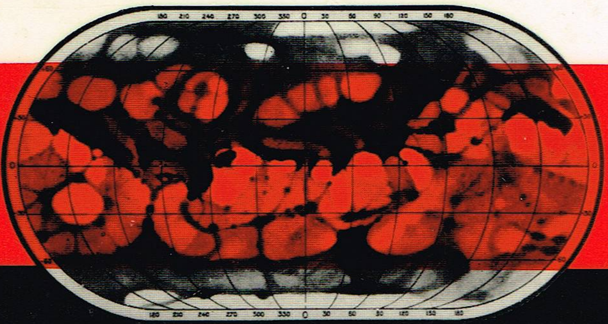
LAROUSSE, DÉPOSITAIRE GÉNÉRAL

Phot. C.E.A. - Mayal et Babcock.









LOUIS LEPRINCE-RINGUET

LA SCIENCE CONTEMPORAINE



les sciences physiques et leurs applications

LAROUSSE

le plus riche dictionnaire
de la langue française

toutes les
connaissances humaines

GRAND LAROUSSE ENCYCLOPÉDIQUE

EN 10 VOLUMES

700 spécialistes internationaux, des moyens techniques ultra-modernes, des méthodes éprouvées garantissent la valeur et la clarté de l'immense documentation contenue dans ce dictionnaire unique au monde.

Le GRAND LAROUSSE ENCYCLOPÉDIQUE en 10 volumes enregistre tous les mots de la langue française (450 000 acceptions de mots), avec l'étymologie, les synonymes, les contraires, des exemples, et la prononciation en alphabet phonétique international.

Dictionnaire du langage, le GRAND LAROUSSE ENCYCLOPÉDIQUE présente également dans l'ordre alphabétique toutes les connaissances, même les plus récentes : les hommes, les œuvres, les événements, les techniques, etc.

Une place très importante a notamment été réservée aux sciences physiques et à leurs applications dans l'industrie : atome, énergie nucléaire, particules subatomiques, accélérateurs de particules, optique électronique, rayons cosmiques, corps semi-conducteurs, mécanique ondulatoire et quantique, isotopes radio-actifs, résines synthétiques, plastiques, etc.

Dans tous les domaines, le GRAND LAROUSSE ENCYCLOPÉDIQUE se révèle un incomparable instrument de travail et de culture.

10 000 pages - 22 000 illustrations et 1 000 cartes en noir - 300 hors-texte et 100 cartes en couleurs.

bibliographie dans chaque volume.

la science
contemporaine

ouvrage en 2 volumes
publié sous la direction de
Louis Leprince-Ringuet
de l'Académie des Sciences
professeur
au Collège de France
et à l'Ecole polytechnique

avec la collaboration de
Jean-Louis Rigal
professeur à la Faculté
des Sciences de Besançon

et de

Maurice Bernard
Jacques Clavier
Louis Mariot
Pierre Mesnage
Francis Muller
Jules Six
Yves Thiry
Jean-Charles Viénot
pour le second volume

la science contemporaine

les sciences physiques et leurs applications

2

LIBRAIRIE LAROUSSE

17, rue du Montparnasse, et boulevard Raspail, 114, Paris

table des matières

UNE RECHERCHE COLLECTIVE



par LOUIS LEPRINCE-RINGUET, professeur au Collège de France et à l'École polytechnique.

LA MÉCANIQUE A L'HEURE SPATIALE



par YVES THIRY, professeur à la Faculté des Sciences de Paris.

LA PHYSIQUE DES ONDES





par PIERRE MESNAGE, professeur à la Faculté des Sciences de Besançon,

et JEAN-CHARLES VIÉNOT, professeur à la Faculté des Sciences de Besançon.



LA RELATIVITÉ

par LOUIS MARIOT, professeur à la Faculté des Sciences de Dijon.



LE MONDE AGITÉ DES PARTICULES ÉLÉMENTAIRES

par JULES SIX, chargé de recherches au C. N. R. S.



UNE INFORMATION TOUJOURS LIMITÉE

par JACQUES CLAVIER, ingénieur de l'Air.



STATISTIQUES ET PROBABILITÉS

par JEAN-LOUIS RIGAL, professeur à la Faculté des Sciences de Besançon.



LA MATIÈRE : ONDES OU CORPUSCULES ?

par FRANCIS MULLER, maître de conférences à l'École polytechnique.



L'ARCHITECTURE DE LA MATIÈRE

par MAURICE BERNARD, maître de conférences à l'École polytechnique.

LES MATHÉMATIQUES DE NOS ENFANTS

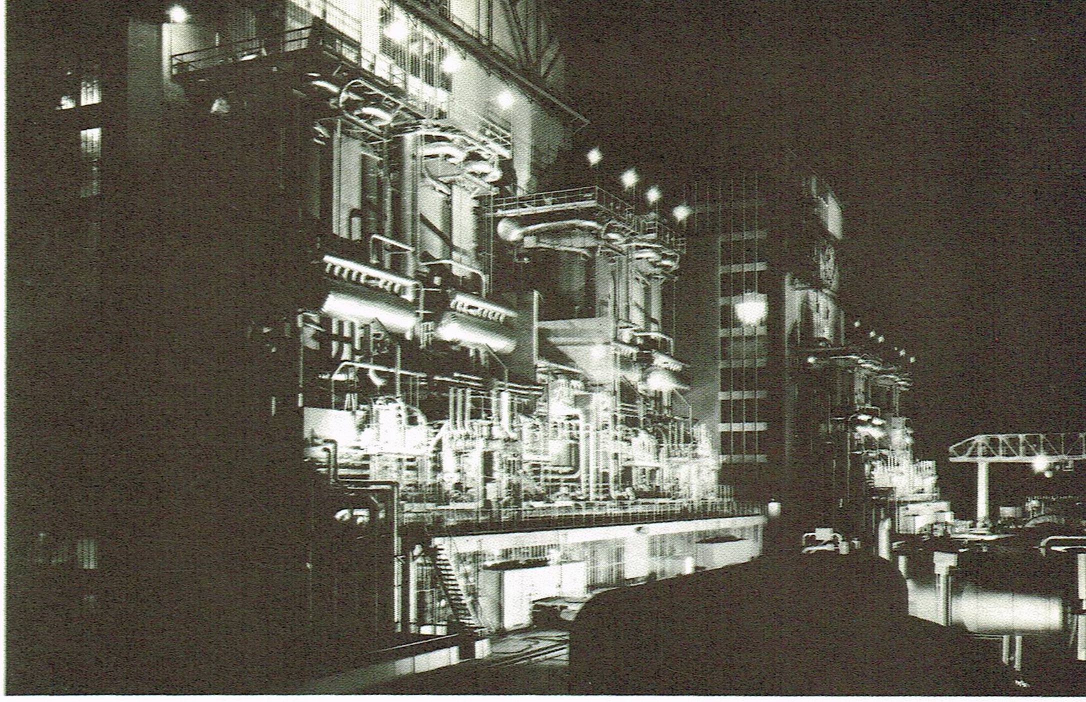
par JEAN-LOUIS RIGAL, professeur à la Faculté des Sciences de Besançon.

COMMUNAUTÉ DES SCIENTIFIQUES

par LOUIS LEPRINCE-RINGUET, professeur au Collège de France et à l'École polytechnique.

pages en couleurs





Au centre de toute recherche : l'homme. L'homme qu'il faut protéger... celui de la pétrochimie et, à Marcoule, celui qui travaille à deux pas du redoutable plutonium.

une recherche collective

Le premier tome de ce gros ouvrage a présenté la richesse extraordinaire du développement des sciences physiques, sans que l'on ait cherché à y séparer science fondamentale et applications. Le premier volet sur l'espace laisse, en effet, une part importante à l'astronomie, à la radio-astronomie, aux rayons cosmiques, ces disciplines qui ne semblent pas comporter d'applications pratiques - ce n'est qu'une apparence -, mais il contient également un large développement sur les fusées et les satellites, fruits d'une science de base bien classique et d'un perfectionnement technologique tout récent et largement évolutif.

Le volet suivant représente le développement d'un ensemble de disciplines spécifiquement modernes : on peut presque dire que c'est l'élaboration d'un langage universel de notre époque. Il s'agit de l'information sous ses diverses formes, de l'optique des lasers, de l'électronique des transistors, des ordinateurs, dont la complexité engendre constamment l'étonnement des plus avertis.

Enfin, le dernier présente trois ensembles puissants à caractère scientifique et industriel à la fois, qui correspondent à trois des principales directions du progrès : hautes températures, macromolécules, énergie atomique. Nous avons voulu marquer bien clairement l'interdépendance de la science de base et de la technique, leur association constante, par l'intermédiaire d'un certain nombre de maillons adaptés, dans l'évolution de chacune de nos grandes disciplines, et, si nous avons coiffé le

Phot. Studio Craven, la Photothèque.

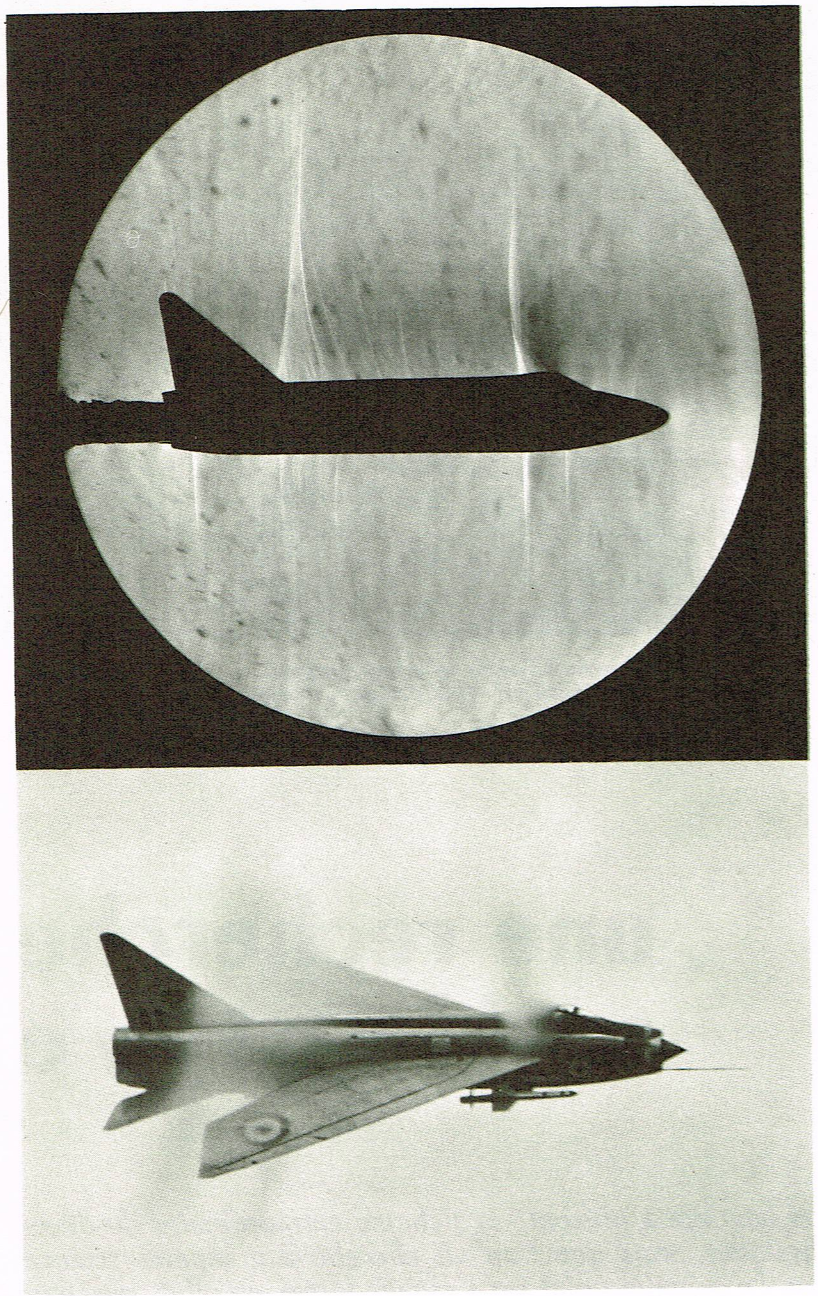
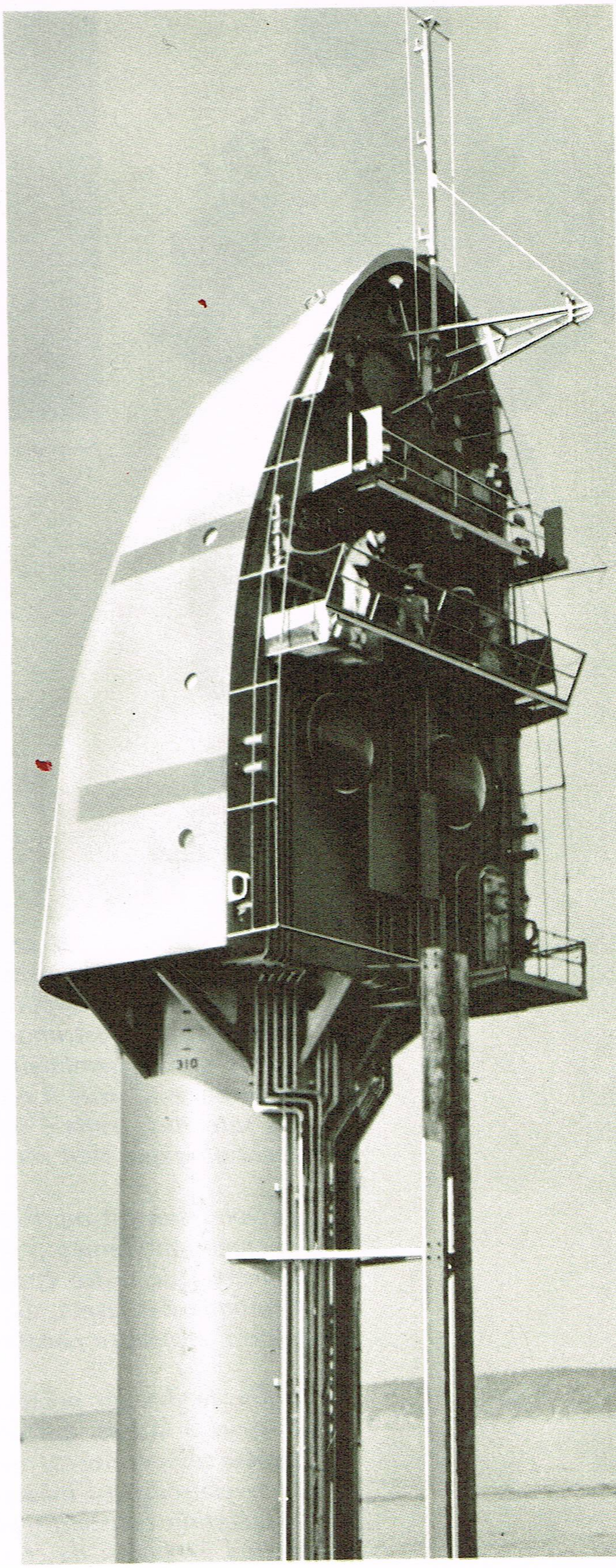


Image extraordinaire que celle des ondes de choc photographiées autour d'un avion en vol supersonique. Exhumée de ses archives par le constructeur, la photographie du haut montre d'autres ondes de choc obtenues lors des essais en soufflerie de la maquette du même avion. La concordance, ainsi établie pour la première fois, est frappante.



L'exploitation rationnelle des ressources que recèle l'Océan implique de vastes programmes de recherches, dont témoigne ce curieux navire qui, une fois arrivé sur place, se dresse verticalement, plongeant au-dessous de l'eau des postes d'observation et conservant la proue au soleil...

Phot. « Shell Aviation », U. S. I. S.

tout par la présentation du CERN et de Nançay, deux grands instituts de science de base - exploration des particules fondamentales par les grands synchrotrons et de l'univers par la radio-astronomie -, c'est précisément pour faire ressentir profondément cette association tout à fait fondamentale : un grand centre de recherches exige autant de développement technologique qu'un grand ensemble industriel à la pointe du progrès.

Mais que de questions laissées sans réponse !
Après ce large tour d'horizon, on s'aperçoit que l'on a saisi, réellement compris, bien peu de choses. Il y a tant de notions de base sous-jacentes lorsqu'on parle d'un transistor, ce minuscule morceau de solide auquel aboutissent quelques fils fins, ou d'une mémoire artificielle, ou d'un plasma, ou d'un noyau d'atome, d'un morceau de plutonium, d'un microscope électronique, que l'on ne peut se satisfaire aisément de simples descriptions truffées d'inconnues, de termes peu accessibles au lecteur non spécialisé.

Ainsi, à côté du premier tome, qui présente le panorama des développements scientifiques, il en faut un autre qui permette de comprendre, d'être finalement satisfait par la découverte ou l'approfondissement des concepts, des lois de la mécanique et de la physique, de tout l'ensemble très fondamental qui constitue notre connaissance de la nature.

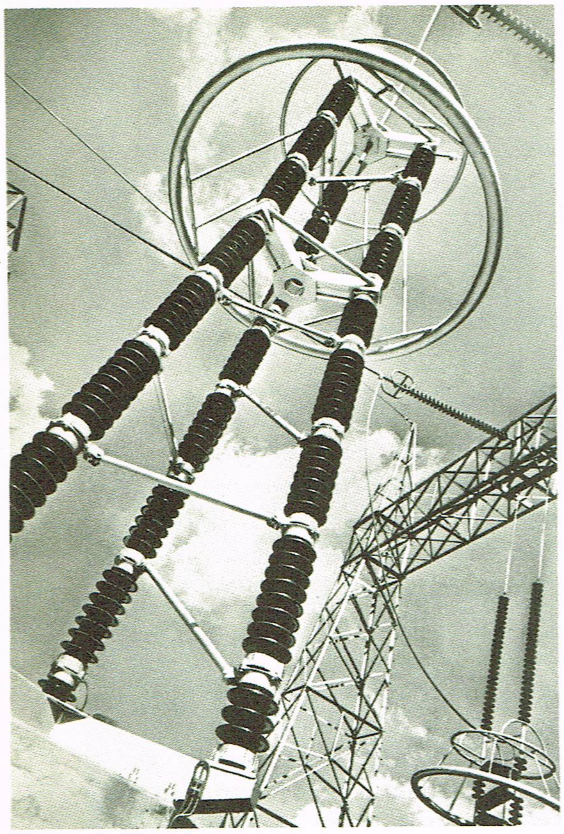
Ainsi le deuxième tome de cet ouvrage apparaît comme indispensable : il expose les bases des sciences physiques.

Mais une telle présentation pour un large public comporte trois écueils redoutables : la légèreté, la difficulté abstraite, l'ennui.

Il est facile et tentant de présenter agréablement un ensemble d'allure scientifique en éludant les problèmes ; avec quelques comparaisons amusantes, on peut " faire passer " les synthèses de Maxwell ou la relativité d'Einstein : c'est une forme de prestidigitation séduisante, ne demandant presque aucun effort au lecteur, mais ne lui apportant qu'une illusion nocive. Que d'explications vulgarisées de la relativité furent décrites par des personnages insuffisamment informés, incapables d'assimiler eux-mêmes l'essentiel des concepts nouveaux, décevants propagateurs des conquêtes magnifiques de l'esprit humain. Il faut de bons scientifiques, bien compétents, pour pouvoir exposer leur science sans confusion, sans tricherie. C'est la première condition. Ainsi, nous avons fait appel à de jeunes professeurs d'université pour les différents chapitres de ce second tome.

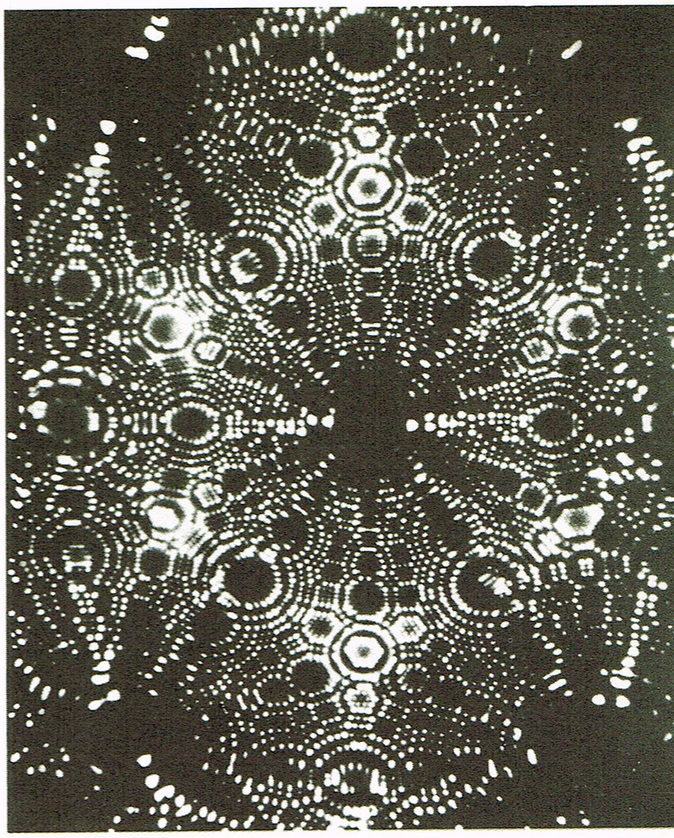
Mais les savants sont souvent redoutables. Leurs exposés sont bons, à coup sûr, mais ils sont durs : on peut leur faire confiance ; on ne les comprend pas toujours. Ils sont trop loin du lecteur moyen et ne peuvent atteindre son niveau. Ils sont habitués à des jeunes gens préparant des examens ou des concours, donc plongés à longueur de journée dans une atmosphère de travail intellectuel et de compétition, rompus aux abstractions et capables, par la fraîcheur de leur esprit, de suivre des raisonnements ardus. Aussi les formules qui semblent naturelles et parfois évidentes aux scientifiques spécialistes recèlent-elles pour le public non initié des difficultés insurmontables ; elles sont loin d'être parlantes pour le lecteur.

Nous avons porté une attention constante à cet aspect du problème. Chaque texte a donné lieu à beaucoup d'allées et venues : entre nous, tout d'abord, pour leur



Sur ce transformateur d'une ligne de 500 000 V, le plus grand parafoudre jamais construit.

Structure, observée en microscopie ionique, de la pointe arrondie d'une aiguille de tungstène grossie 2 millions de fois.

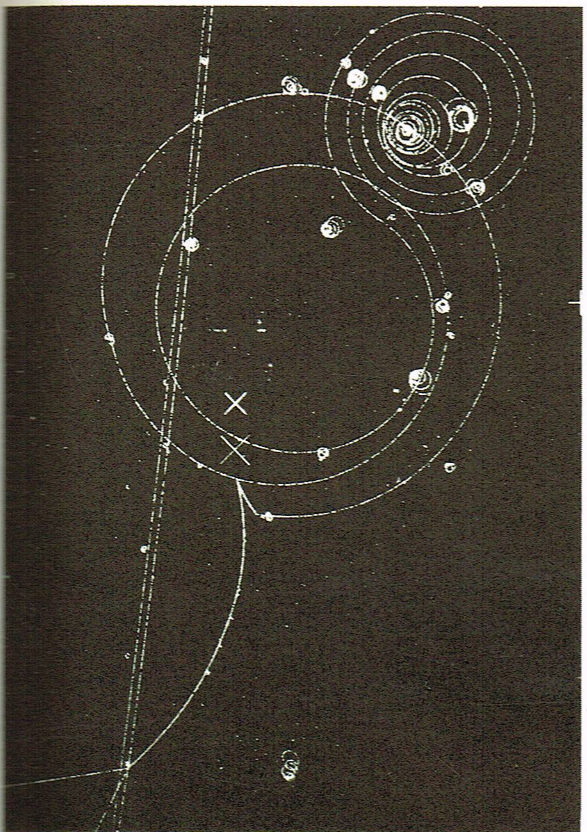


harmonisation, mais aussi entre les auteurs et des lecteurs peu initiés. Cela n'a pas été sans difficulté, car la science est difficile et nous sommes, nous-mêmes, bien loin de tout comprendre ; même la très classique théorie de Maxwell, englobant la lumière dans l'ensemble des phénomènes électromagnétiques, cette théorie presque centenaire qui régit la propagation de la radio, de la télévision, de la lumière, des radars, le fonctionnement des guides d'onde, n'est pas encore très facile à assimiler, et le "courant de déplacement", dont il faut bien parler, reste encore un peu mystérieux.

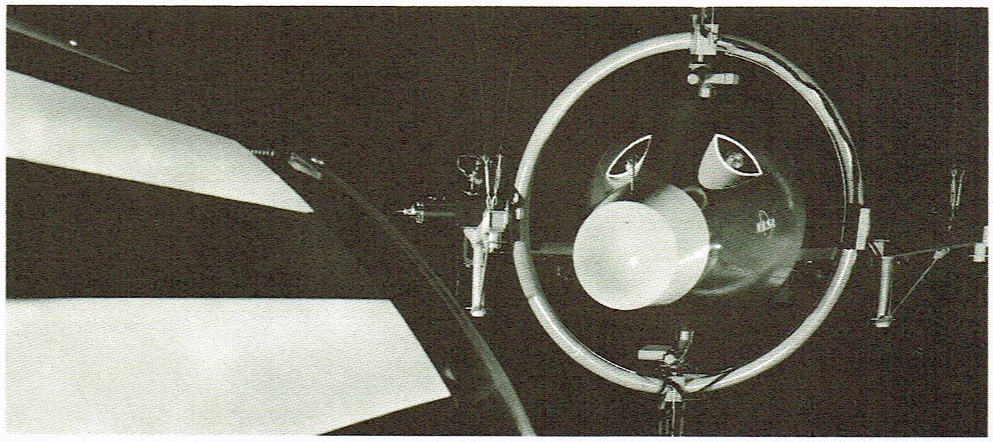
D'ailleurs, si les phénomènes eux-mêmes - rayons X, ultraviolet, résonances, voire collisions nucléaires - peuvent se décrire aisément, il n'en est pas de même des grandes synthèses, des théories de la relativité et de la mécanique ondulatoire. Ces théories cherchent à expliquer les réalités objectives, celles dont il faut obligatoirement tenir compte pour les grouper dans une explication cohérente. Elles ont été élaborées progressivement, elles s'épaulent et se complètent sans se détruire, elles s'adaptent de mieux en mieux aux phénomènes connus, elles sont de plus en plus générales et satisfaisantes. Ainsi la théorie de Maxwell ne visait que l'ensemble des phénomènes ondulatoires. La relativité étend cela aux très grandes vitesses et aux systèmes en mouvement relatif. Puis la première théorie quantique, celle de Bohr-Sommerfeld, essaie d'expliquer les phénomènes d'allure corpusculaire et l'émission, par les atomes, de raies lumineuses. A vrai dire, c'est une théorie de début, cherchant plutôt les règles du jeu qu'une explication cohérente : à ce degré, elle n'est pas satisfaisante, et Bohr lui-même l'a bien compris.

Alors on a cherché à expliquer les phénomènes de l'échelle de l'atome, et ce furent les magnifiques efforts qui ont permis de construire la mécanique ondulatoire, et, plus généralement, la mécanique quantique. C'est une construction difficile : il a fallu abandonner bien des notions classiques, et c'est difficile d'abandonner des

Phot. U. S. I. S.



Florilège des particules dans une chambre à bulles de Saclay-École polytechnique. Des deux particules de haute énergie venant du haut, une donne lieu à plusieurs interactions, et la trace qui remonte est due successivement à un méson π lent, un méson μ, un électron positif (spirale à gauche) et un électron négatif (spirale à droite).



Dès maintenant, les astronautes mettent au point les techniques du rendez-vous spatial : la capsule, qui revient d'un astre, s'apprête à s'accrocher à l'astronef resté sur une orbite d'attente.

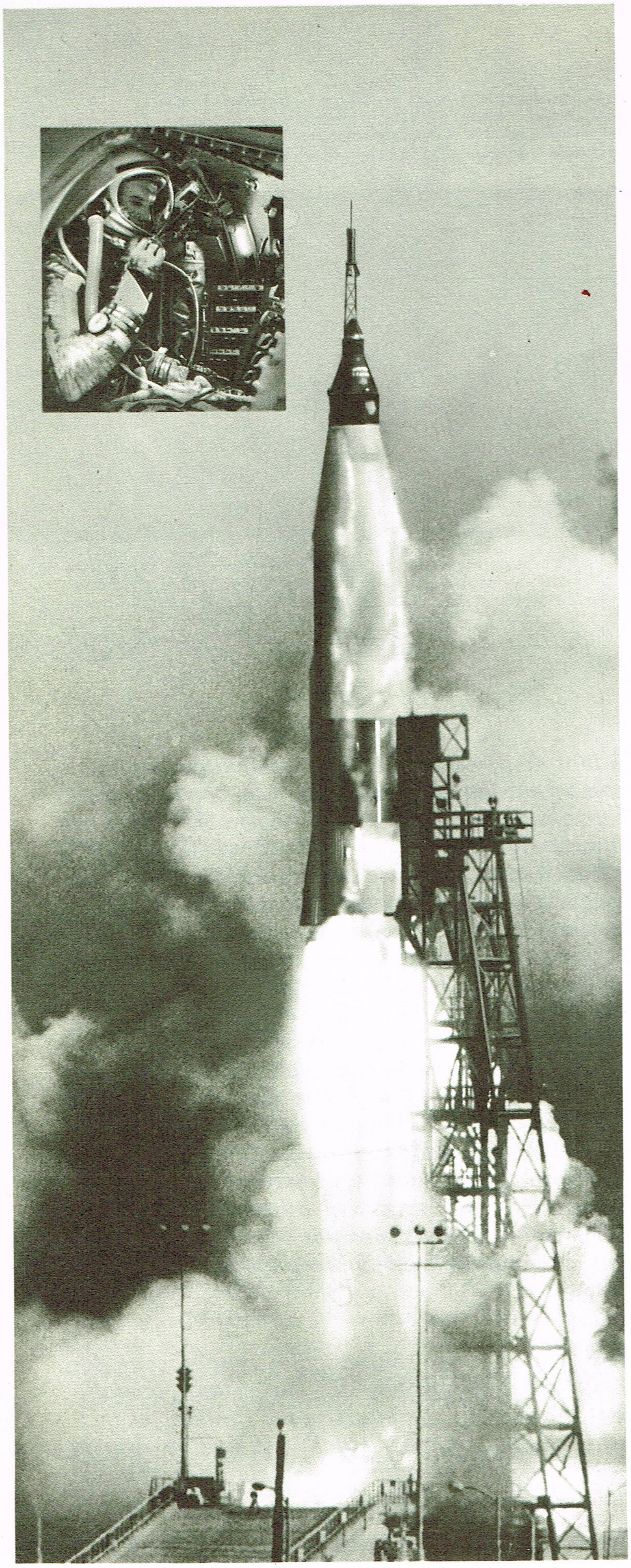
concepts familiers. Nous avons déjà tant de peine à modifier nos habitudes au cours de notre existence ! Mais quel succès ! Les spectres des atomes apparaissent alors naturels, découlent de l'étonnant formalisme introduit par les théories nouvelles. La stabilité des atomes, la radio-activité, la création des particules, les antiparticules, et, même, de façon plus générale, la cohésion des molécules et les propriétés des solides. Bref, non seulement la physique de l'infiniment petit, mais toute la chimie se construit alors de façon satisfaisante.

Nous n'avons pas voulu éluder cela et nous avons traité la relativité d'Einstein et la mécanique quantique à laquelle sont attachés les noms de Louis de Broglie, Schrödinger, Heisenberg, Dirac, Pauli. Peut-être est-ce là une gageure, mais nous avons pensé que ce serait malhonnête que de ne pas essayer de donner un aperçu, aussi simple, élémentaire et clair que possible, des grandes synthèses permettant de comprendre notre univers et son évolution. Que le lecteur ne s'y trompe pas, néanmoins : il lui faudra faire un effort dont nous préciserons le sens en tête des chapitres sur la physique quantique. Ainsi, la lecture de cet ouvrage présentera des passages faciles et d'autres plus difficiles. Il en est ainsi de tout dans l'existence ; les montagnards qui tentent une ascension savent bien que certains passages seront aisés et que d'autres demanderont toute leur attention active : on ne parvient au but que par une telle succession.

Enfin, nous avons cherché à éviter l'ennui : des résumés, des redites sous des formes diverses, des schémas aussi parlants que possible, des illustrations scientifiques, des images d'accompagnement relanceront, espérons-nous, l'intérêt et faciliteront les traversées difficiles. Notre souhait le plus ardent est que la grande magnificence de la science moderne, unissant les grandioses réalisations expérimentales aux audacieuses théories, apparaisse finalement dans son éclat au lecteur assidu.

Louis LEPRINCE-RINGUET

Phot. École polytechnique, U. S. I. S.



LA MÉCANIQUE

Jusqu'à ces dernières années, l'homme a été rivé à la surface de la Terre, qu'il n'a pu survoler, ou pénétrer, que de quelques kilomètres. Il n'a eu à sa disposition qu'un domaine pratiquement à deux dimensions, mince couche sphérique à la surface d'un petit astre de l'univers.

Malgré cela, il a su, au cours des siècles, amener à un haut degré de perfection les sciences et les techniques. Celles-ci lui ont permis, en particulier, de briser les chaînes qui l'attachent à la Terre et de partir à la conquête de l'espace. L'importance de cette conquête est telle que l'on a pu dire que nous vivons actuellement un nouvel âge de l'humanité : l'ère spatiale.

La science du mouvement, la mécanique, joue naturellement un rôle important dans cette conquête, qui n'a été possible qu'après que l'humanité eut su acquérir une compréhension correcte des lois de la mécanique. Apporté par le développement passé des sciences et des techniques, l'avènement de l'ère spatiale concourt dès à présent puissamment à leur développement futur. Si la mécanique classique n'a subi aucun bouleversement dans ses principes mêmes, les observations et l'analyse du mouvement des satellites artificiels ont apporté des connaissances nouvelles de très grande importance : l'espace environnant la Terre à des altitudes qui n'avaient encore jamais été atteintes, la Terre elle-même, sa forme, la répartition de ses masses sont de mieux en mieux connus.

Les mouvements des astres avaient été compris par Copernic, expliqués par Newton, mis en équation par Lagrange et Laplace, mais n'intéressaient plus que de rares spécialistes ; la mécanique céleste — mère de la dynamique et aussi de la plupart de nos connaissances mathématiques — semblait bien passée de mode, lorsque le vol des engins spatiaux a suscité un renouveau d'intérêt pour elle. Les problèmes nouveaux posés aux mécaniciens les ont obligés à rénover ou à adapter leurs

Phot. U. S. I. S.

A L'HEURE SPATIALE

par
Yves THIRY

méthodes traditionnelles et à développer des méthodes originales de calcul.

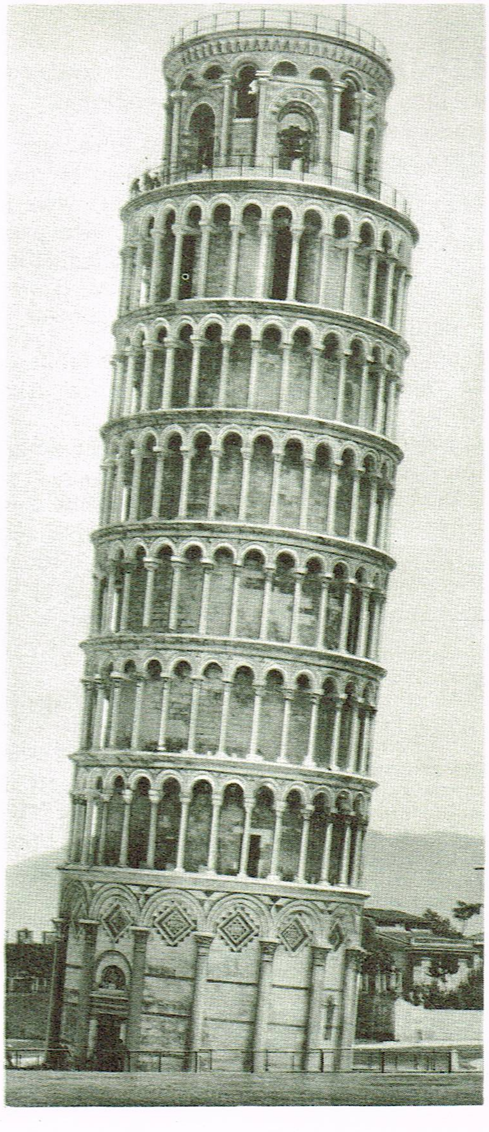
Qui ne veut à présent comprendre le mouvement des engins spatiaux, discuter de la difficulté d'un « tir à la Lune » ou d'un atterrissage sur Mars ?

Or, les réponses à des questions de mécanique, souvent d'apparence bien anodine, nécessitent fréquemment une analyse assez fine et une compréhension assez poussée des lois qui régissent cette science.

Présentons un peu au hasard *(ci-contre et p. 9)* quelques questions de mécanique, les unes relatives à des phénomènes couramment observés à la surface de la Terre, les autres relatives à la préparation ou à la réalisation du premier voyage d'un homme dans l'espace. Et d'abord ces différentes questions sont-elles de natures différentes, ou se rattachent-elles aux mêmes principes ? Autrement dit, y a-t-il une seule théorie mécanique pour expliquer les diverses catégories de mouvements auxquelles on peut penser ?

Voilà des questions devant lesquelles certains lecteurs hésiteront peut-être. Elles sont de pure mécanique ; elles n'exigent certainement aucune connaissance approfondie de la constitution ultime de la matière, et il semble évident que l'on doive pouvoir y répondre sans être un puits de science en mathématiques. Pourquoi donc, alors, certains hésiteront-ils, ou, ce qui est plus grave, donneront-ils d'emblée et de toute bonne foi une réponse erronée ?

C'est que la mécanique, qui devait apparaître comme le premier chapitre de la physique, le plus simple et le plus directement accessible, est en fait une science complexe. Pour s'en convaincre, le mieux serait de revenir un peu à son histoire. Aujourd'hui, cette science a été amenée à un haut degré de perfection, et elle constitue une théorie cohérente assise sur des bases solides, qui apporte des solutions satisfaisantes dans



Des deux sphères de même rayon, l'une en bois, l'autre en fonte, que Galilée lâcha du haut de la tour de Pise, laquelle est arrivée la première au sol ?

Phot. Brogi.

un domaine certes restreint, mais d'une importance scientifique et technique considérable.

Mais l'homme est mal placé, entraîné qu'il est dans le mouvement compliqué de la Terre, pour comprendre les lois de la mécanique. Il n'y a que trois siècles environ, après avoir vécu des millénaires de recherches, de tâtonnements, de lueurs fugitives et de retours à l'obscurité, qu'il lui fut donné, dans la personne de Newton, de pouvoir énoncer les principes qui devaient servir de base à cette science (1687). Mais, d'autre part, ce fut juste au moment où la mécanique classique fut complètement élaborée que les grandes découvertes de la fin du XIXe siècle et du début du XXe firent éclater son cadre trop étroit. L'étude des mouvements, en relation avec les phénomènes nouvellement mis en évidence, put de moins en moins être dissociée des autres phénomènes physiques, et deux nouvelles mécaniques, la mécanique ondulatoire et la mécanique relativiste, virent le jour. L'histoire de la mécanique classique est donc très différente de celle des autres sciences : des millénaires d'obscurité, trois siècles de splendeur, un échec spectaculaire en face de phénomènes qui n'étaient manifestement pas de son domaine, un regain d'actualité dû au lancement des engins spatiaux. Et cela pour une science qui n'avait pas cessé d'être à la base de la presque totalité de nos réalisations techniques.

Mais ce n'est pas de considérations historiques que nous voulons entretenir le lecteur. La mécanique, dans son état actuel, est une science bien vivante et fondamentale, et c'est cet état actuel que nous avons l'intention de présenter.

Retenons donc simplement de notre brève allusion à l'histoire de la mécanique que l'homme est mal placé pour comprendre ses lois et qu'il lui a été difficile de situer les mouvements dans le cadre qui leur convient.

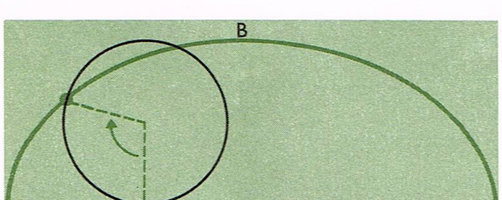
les accélérations déterminent les mouvements

Assis à mon bureau, j'ai le sentiment très net de mon immobilité, mais j'éprouve exactement la même sensation dans une « Caravelle » qui m'emmène à 800 km à l'heure, ou lorsque je suis confortablement installé dans une couchette du Simplon-Orient-Express, traversant la Brie à sa vitesse de croisière. Ou, plutôt, seuls le bruit rythmé des roues sur les joints de dilatation ou le défilement des lumières ou du paysage devant les vitres peuvent alors me rappeler que je voyage; ou, encore, lorsque le convoi prend de la vitesse, ralentit ou aborde une courbe, les accélérations que je subis, et qui me plaquent sur la banquette ou me font perdre l'équilibre.

Lorsqu'en gare de Lyon le rapide démarre, et que commencent à défiler sous mes yeux les vitres du train situé sur la voie parallèle, je suis incapable de dire lequel des deux vient de partir, tant sont douces les accélérations des grands convois internationaux.

En automobile, le sentiment que j'ai de ma vitesse dépend essentiellement de l'état de la route. Sur les tronçons d'excellentes routes que nous commençons à connaître, je suis obligé de jeter de fréquents coups d'œil à mon tableau de bord si je veux, pour des raisons de sécurité, de rodage ou de consommation, effectuer mon voyage sans dépasser la vitesse que je me suis fixée.

Un cycliste se déplace sur la route d'un mouvement rectiligne uniforme. La valve de la roue avant a, par rapport au référentiel constitué par le cadre de la bicyclette, un mouvement simple de rotation uniforme. La trajectoire de la valve par rapport au sol est une *cycloïde*. Par rapport au sol, la valve est presque immobile en A; en B, au contraire, elle va deux fois plus vite que la bicyclette.



quand donc suis-je en mouvement ?

Pour répondre à cette question, je suis obligé de me placer, de me *repérer*, par rapport aux objets qui m'entourent : dans mon bureau, je suis immobile lorsque les distances qui me séparent de ces objets ne varient pas. Assis dans mon compartiment, je suis immobile par rapport au train, alors même que je me rapproche de la gare que je m'étais fixée comme but; de même, encore, lorsque je quitte ma place pour gagner la voiture-restaurant, j'ai une idée très nette de mon déplacement et de son but.

par rapport à quoi suis-je en mouvement ?

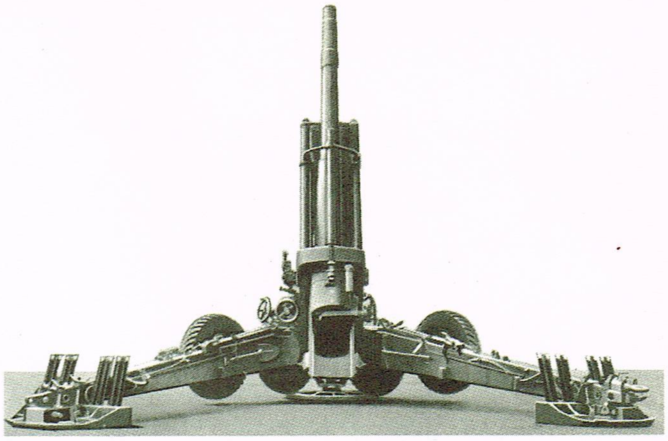
Ainsi donc, ce n'est qu'en évaluant à chaque instant les distances à différents objets que l'on peut déceler l'existence d'un mouvement. Naturellement, ces objets ne doivent pas être choisis n'importe comment : aucune conclusion nette ne peut être tirée de l'évaluation de mes distances par rapport à des objets dont les distances mutuelles varient.

Mais il est des corps pour lesquels la distance entre deux quelconques de leurs points ne varie pas; ce sont les corps solides. Ainsi, nous ne pouvons parler de repos ou de mouvement que par rapport à un certain solide : le solide de référence. On dit que le mouvement est une notion relative, et l'on parle de la **relativité** du mouvement. Les mots « mouvement », « trajectoire », « vitesse », « accélération » n'ont aucun sens si le solide de référence n'est pas précisé.

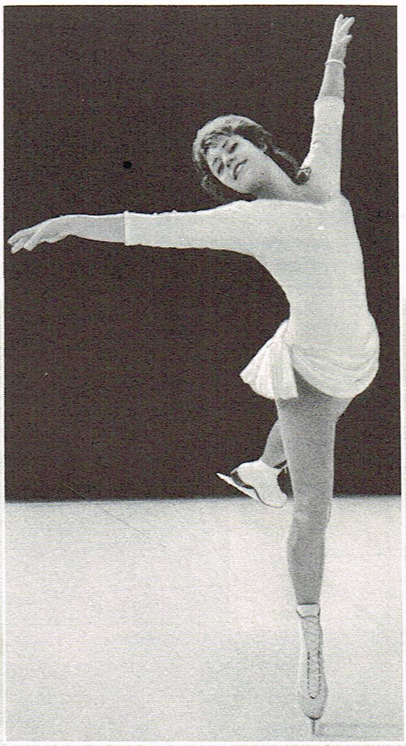
Le solide de référence n'est pas, en général, constitué par un seul corps indéformable tel qu'une maison ou un wagon de chemin de fer. Nous aurons, par exemple, à envisager le solide stellaire, en admettant que, si l'on ne s'intéresse pas à des durées trop considérables, les distances entre les différentes étoiles ne varient pas. Ce solide est schématisé par la sphère céleste, dont il faudra préciser le centre, et sur laquelle les étoiles fixes dessinent la figure apparemment invariable des constellations.



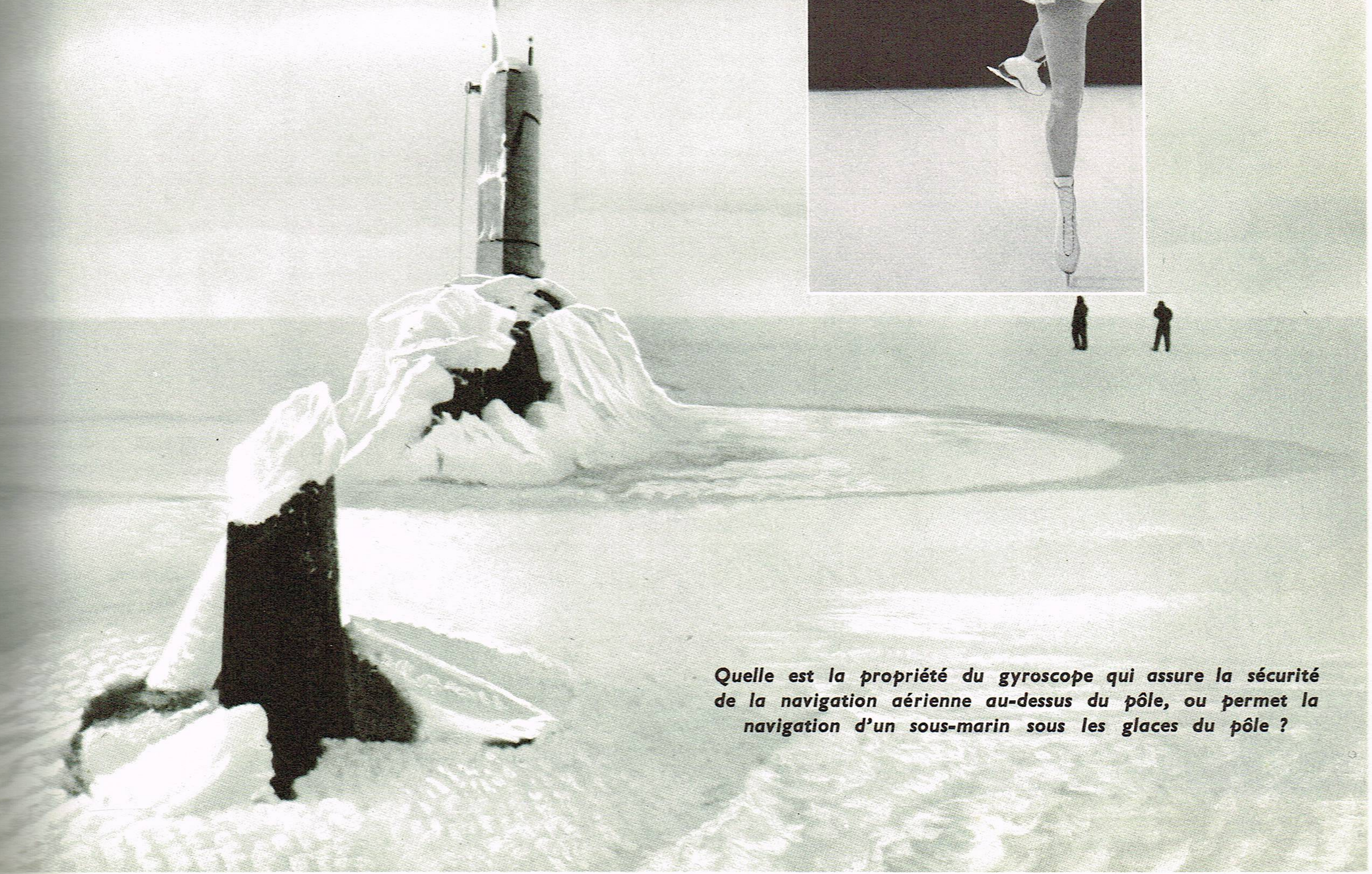
Comment s'explique l'état d'apesanteur où se trouve un cosmonaute ?



Le projectile vient de partir : lancé rigoureusement suivant la verticale, retombera-t-il dans la bouche à feu ou plus à l'ouest, s'il faut tenir compte de la dérobade sous ce projectile de la Terre dans sa rotation vers l'est ?

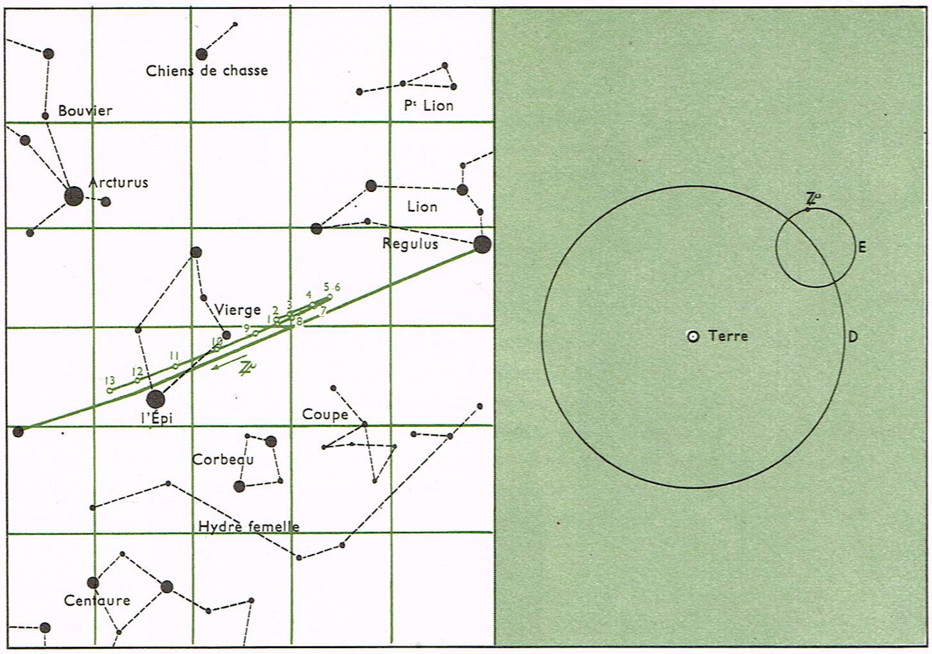


Comment la patineuse, arrêtée sur la pointe de son patin, peut-elle à volonté accélérer ou ralentir sa rotation ?

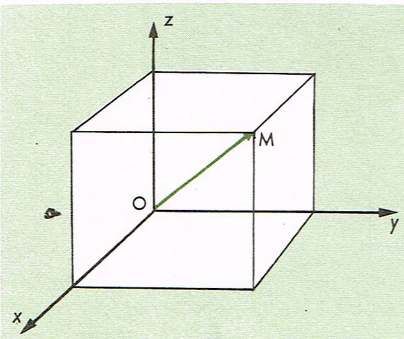


Quelle est la propriété du gyroscope qui assure la sécurité de la navigation aérienne au-dessus du pôle, ou permet la navigation d'un sous-marin sous les glaces du pôle ?

Phot. U. S. I. S., Presse-Sports.

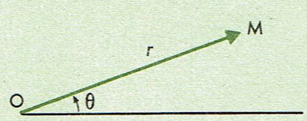


Les mouvements des planètes, astres errants, en projection sur la sphère céleste, ont été décelés dès la plus haute Antiquité. La carte ci-dessus indique les positions de Jupiter (♃) aux premiers jours de chaque mois de l'année 1957.
***A droite,* la description de Ptolémée (IIe s. apr. J.-C.) rappelle celle du mouvement de la valve de la bicyclette : dans le plan de l'écliptique, Jupiter décrit l'*épicycle* E dont le centre se déplace sur le déférent D.**



coordonnées cartésiennes

Le trièdre trirectangle O *x y z*, conçu comme constitué de particules appartenant au solide référentiel, matérialise le référentiel. Les trois distances (algébriques) du mobile M aux trois faces de ce trièdre sont les trois coordonnées cartésiennes *x*, *y*, *z* de M. Ce sont aussi les trois *composantes* du vecteur $\overrightarrow{OM}$ (segment OM décrit de O vers M) sur les axes de ce trièdre.

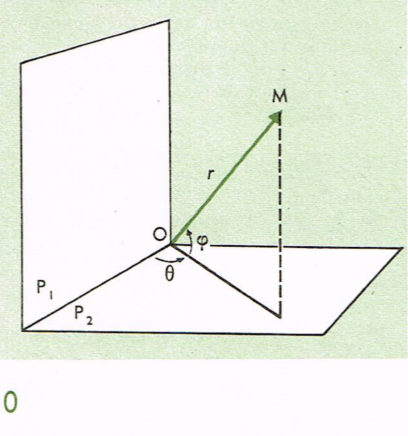


coordonnées polaires

Si le mobile se déplace dans un plan appartenant au référentiel les deux coordonnées polaires : *r*, θ, suffisent.

théodolite

Lors d'une visée, on mesurera le déplacement du plan vertical de la lunette, puis le déplacement de la lunette dans son plan vertical; il faudra, de plus, mesurer la distance du mobile à l'aide d'un télémètre.



coordonnées sphériques

Les plans P1 et P2 appartiennent au référentiel. Les coordonnées sphériques de M sont *r*, θ, φ.

Abandonnons donc l'expression « solide de référence » et remplaçons-la par un mot nouveau : **référentiel.**

Dans le langage courant, le référentiel est souvent sous-entendu. Le mouvement étant décelé par le fait que les distances du mobile aux différents points de la Terre varient au cours du temps, c'est la Terre qui constitue notre référentiel ordinaire. Mais on peut être amené à considérer d'autres repères (le train ou la voiture qui nous véhicule, etc.), qu'il importe de préciser exactement.

le repérage d'un mobile

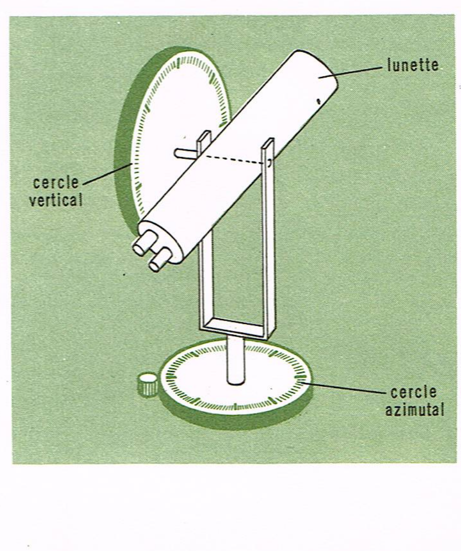
Le problème du repérage d'un mobile par rapport à un certain référentiel est celui du rendez-vous : les questions qu'il faut poser sont « *qui?* », « *où?* » et « *quand?* ».

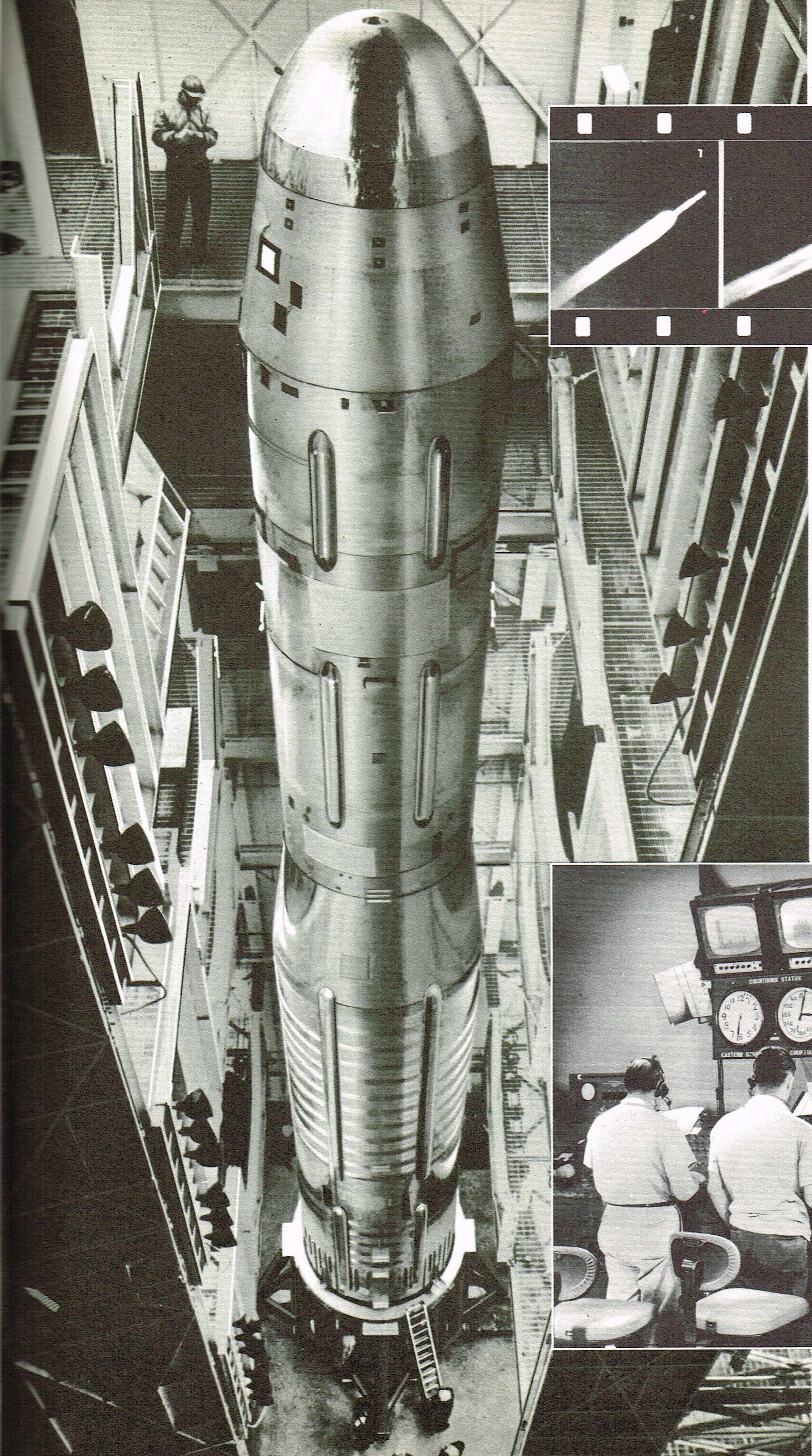
qui ?

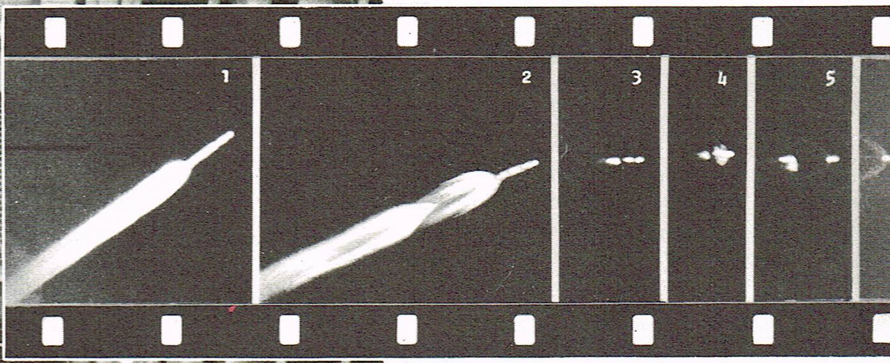
Le mobile est caractérisé à l'aide d'un des procédés ordinaires d'identification : on parlera, par exemple, de la voiture immatriculée 7341 LL 75, de l'avion d'Air France effectuant le vol 2482 du 3 mai, de 1959 θ (« Lunik III », qui a photographié la face cachée de la Lune).

où ?

La position du mobile à chaque instant pourra être repérée, non seulement comme nous l'avons vu par des mesures de distances à des points du référentiel, mais aussi par des mesures d'angles. Trois mesures sont nécessaires et suffisantes pour repérer la position d'un point dans l'espace. Les nombres obtenus s'appellent les **coordonnées** du mobile par rapport au référentiel.



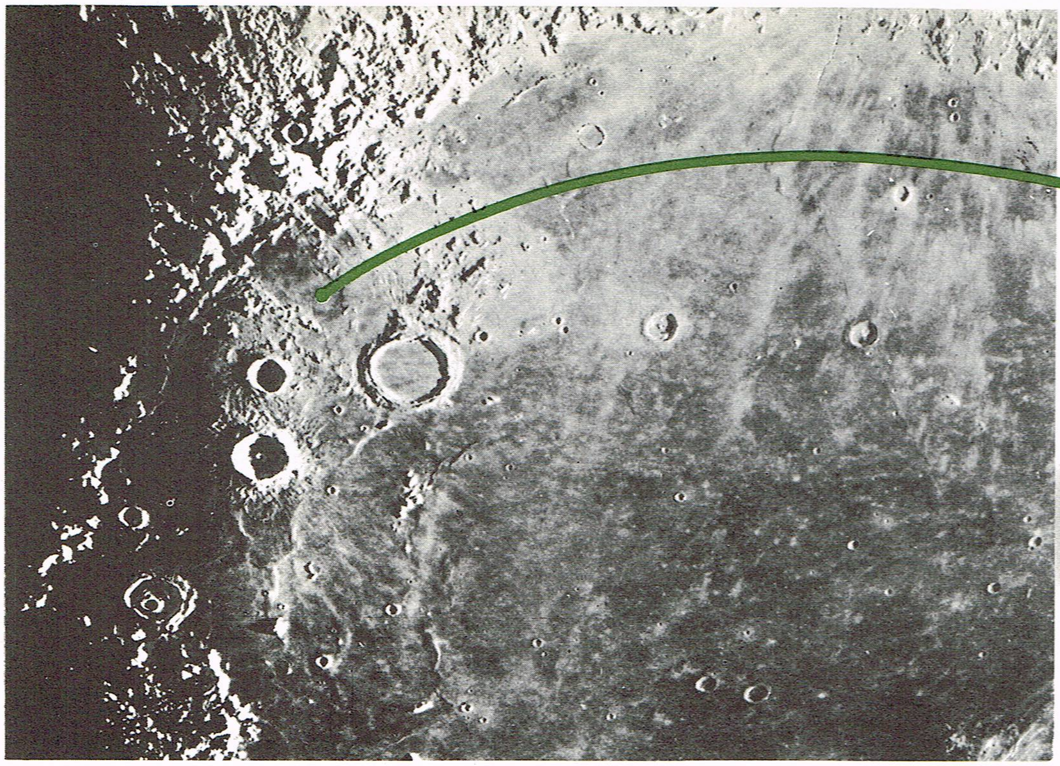


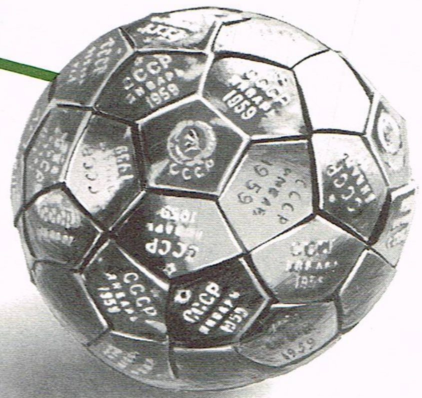


qui ? où ? quand ?

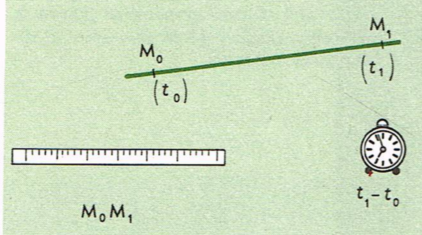


Phot. U. S. I. S.



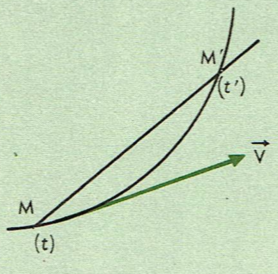


Un rendez-vous spectaculaire : celui de « Lunik II » avec la Lune, le 13 janvier 1959, quelques minutes après 21 heures.

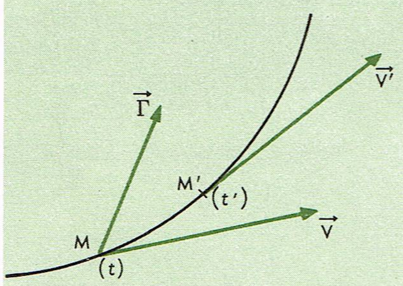


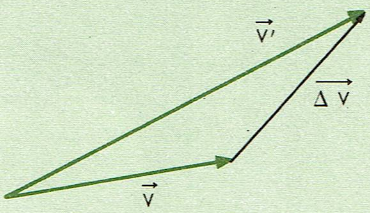
$$\vec{V} = \frac{\overrightarrow{M_0M_1}}{t_1 - t_0} = \frac{\overrightarrow{\Delta M}}{\Delta t}$$

La vitesse moyenne du mobile M entre les instants t_0 et t_1 est le rapport entre la *variation* vectorielle $\overrightarrow{\Delta M}$ de sa position à l'*accroissement* Δt du temps.



Lorsqu'on considère des instants t' de plus en plus proches de l'instant t, la vitesse moyenne $\frac{\overrightarrow{MM'}}{t'-t}$ tend vers la *vitesse vraie* $\vec{V}$ à l'instant t. C'est la limite de $\frac{\overrightarrow{\Delta M}}{\Delta t}$, que l'on note $\frac{dM}{dt}$ et qui est la *dérivée* du point M par rapport au temps.





Pour obtenir l'accélération $\vec{\Gamma}$ du mobile M à l'instant t, il faudra considérer l'accroissement $\overrightarrow{\Delta V}$ de sa vitesse pendant l'intervalle de temps Δt, diviser cet accroissement par Δt et passer à la limite : $\vec{\Gamma} = \lim \frac{\overrightarrow{\Delta V}}{dt} = \frac{d\vec{V}}{dt}$, que l'on note $\frac{d^2M}{dt^2}$, est la dérivée de $\vec{V}$ ou la *dérivée seconde* du point M par rapport au temps.

quand ?

La détermination de l'instant auquel a été fait le repérage de position du mobile pose la délicate question de la mesure du temps. N'entrons pas dans les détails et adressons-nous aux spécialistes. Une des missions des astronomes est précisément de construire et de diffuser le temps. Ils utilisent comme horloge le système solaire, ou plus particulièrement un phénomène céleste susceptible d'observations très fréquentes et très précises, par exemple l'occultation des étoiles par le disque lunaire. Ils construisent alors le temps et nous diffusent l'heure par l'horloge parlante ou par des signaux horaires transmis par radio. Nous leur faisons confiance et nous réglons nos montres selon leurs indications. Ce temps nous permet de fixer l'heure des rendez-vous.

vitesse et accélération

La cinématique est cette partie de la mécanique qui traite des trajectoires des mobiles et de la façon dont ces trajectoires sont décrites au cours du temps. Il ne suffit pas de savoir repérer la position du mobile à un instant déterminé, il faut savoir étudier l'évolution de cette position au cours du temps. Les grandeurs physiques qu'il suffit de considérer pour rendre compte de cette évolution sont la vitesse et l'accélération.

Tout le monde parle de vitesse. On a même pu dire de notre civilisation que c'était la civilisation de la vitesse : aller vite, c'est faire beaucoup de choses en

Phot. Observatoire du Mont-Wilson, Keystone.

peu de temps. Bien entendu, une telle phrase est beaucoup trop vague pour constituer une définition, et il faut la préciser. On retrouvera la notion de vecteur, déjà aperçue lors de l'étude des coordonnées cartésiennes et des problèmes de localisation d'un point dans l'espace, et on verra apparaître le vecteur vitesse et le vecteur accélération. La direction et le sens de la vitesse demandent à être connus à chaque instant au même titre que son intensité (c'est-à-dire sa mesure, en mètres à la seconde ou en kilomètres à l'heure).

On a déjà vu apparaître l'accélération, seule ressentie pendant le mouvement. Nous verrons le rôle essentiel qu'elle joue en mécanique, directement liée qu'elle est aux forces; c'est également une grandeur vectorielle attachée à ce mobile qui indique comment varie sa vitesse.

mouvements rectilignes

La position du mobile M est donnée à chaque instant par son abscisse $x = OM$ qui est une fonction $x(t)$ du temps.

Le vecteur vitesse est porté par la trajectoire Ox, il a pour mesure algébrique sur cet axe :

$$v = \lim \frac{x(t') - x(t)}{t' - t} = \lim \frac{\Delta x}{\Delta t} = \frac{dx}{dt}$$; c'est la dérivée de la fonction $x(t)$.



Le vecteur accélération est également porté par Ox, il a pour mesure algébrique sur cet axe :

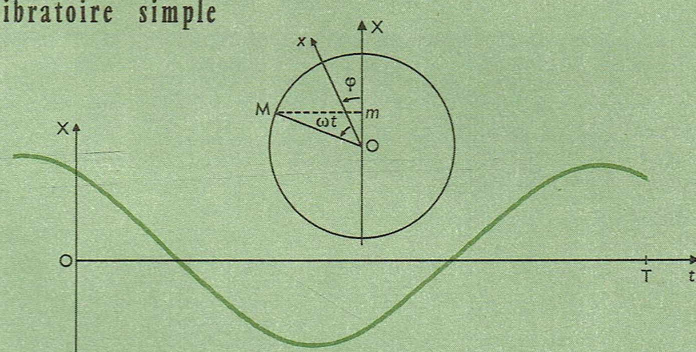
$$\gamma = \lim \frac{v(t') - v(t)}{t' - t} = \lim \frac{\Delta v}{\Delta t} = \frac{dv}{dt} = \frac{d^2x}{dt^2}$$; c'est la dérivée seconde de la fonction $x(t)$.

Dans ce cas, particulièrement simple, le mouvement est accéléré, décéléré ou uniforme, suivant que la grandeur de la vitesse augmente, diminue ou reste constante.

mouvement vibratoire simple

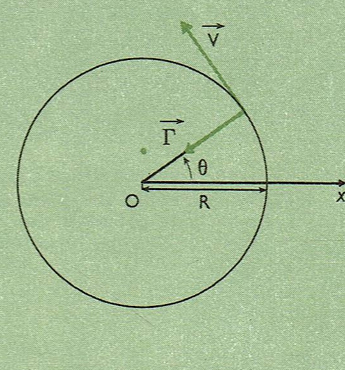
C'est la projection sur un diamètre d'un mouvement circulaire uniforme. La position du mobile m est donnée à chaque instant par son abscisse $X = R \cos(\omega t + \varphi)$. Sa vitesse est la projection sur OX de la vitesse de M; elle a pour mesure algébrique sur OX, $- R\omega \sin(\omega t + \varphi)$. Son accélération est la projection sur OX de l'accélération de M; elle a pour mesure algébrique sur OX : $- \omega^2 X$: elle est dirigée vers O et est proportionnelle à l'élongation Om.

ω est la pulsation; $T = 2\pi/\omega$ la période; $\nu = 1/T$ la fréquence; φ le déphasage.

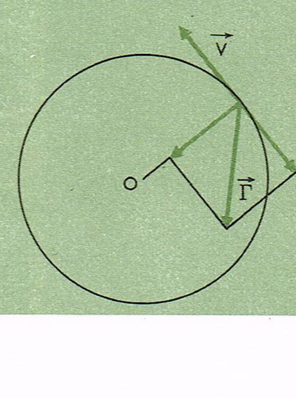


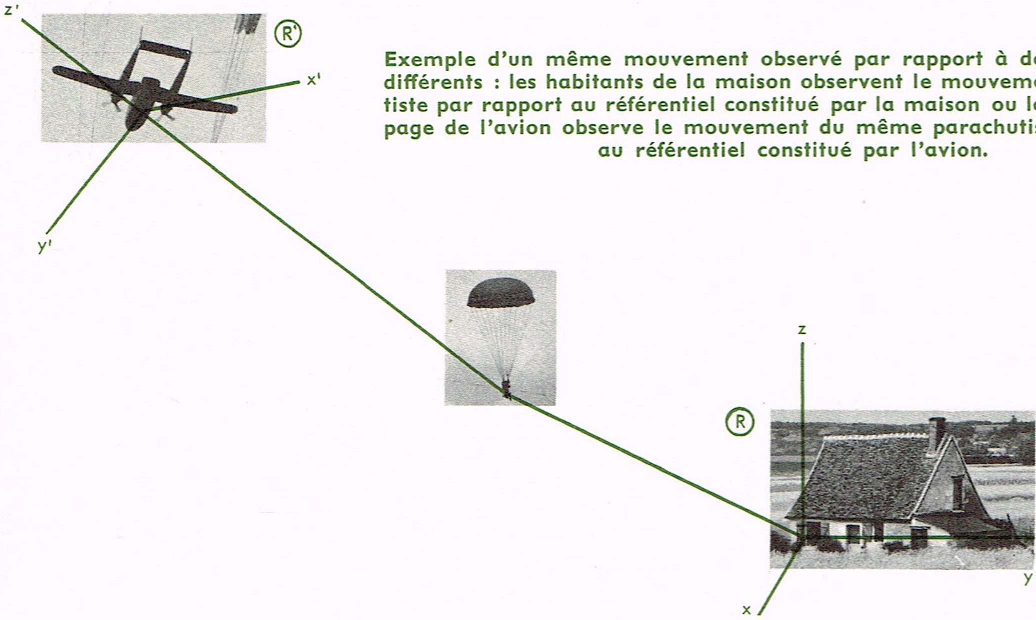
mouvements circulaires

Revenons au mouvement circulaire de la valve de la roue du cycliste page 8. La position de la valve est repérée à chaque instant par l'angle θ; cet angle est proportionnel au temps $\theta = \omega t$ (le coefficient de proportionnalité ω est dit « vitesse angulaire de la valve »). Le vecteur-vitesse de la valve est tangent à la trajectoire, dans le sens du mouvement, et a pour grandeur $R\omega$. Cette grandeur est constante et l'accélération n'a à rendre compte que de la variation de la direction de $\vec{V}$; le vecteur accélération $\vec{\Gamma}$ est porté par le rayon dirigé vers le centre (l'accélération est centripète) et a pour grandeur $R\omega^2$. Le mouvement est périodique, de période $T = \frac{2\pi R}{v} = \frac{2\pi}{\omega}$: le mobile repasse à la même position et avec la même vitesse au bout du temps T.



Pour un mouvement circulaire non uniforme, l'angle θ de repérage est une fonction quelconque du temps. La vitesse angulaire $\omega = \frac{d\theta}{dt}$ varie avec le temps, ainsi que la grandeur $R\omega$ de V. En plus de l'accélération centripète de grandeur $R\omega^2$, il existe une accélération tangentielle, de mesure $R\frac{d\omega}{dt}$.



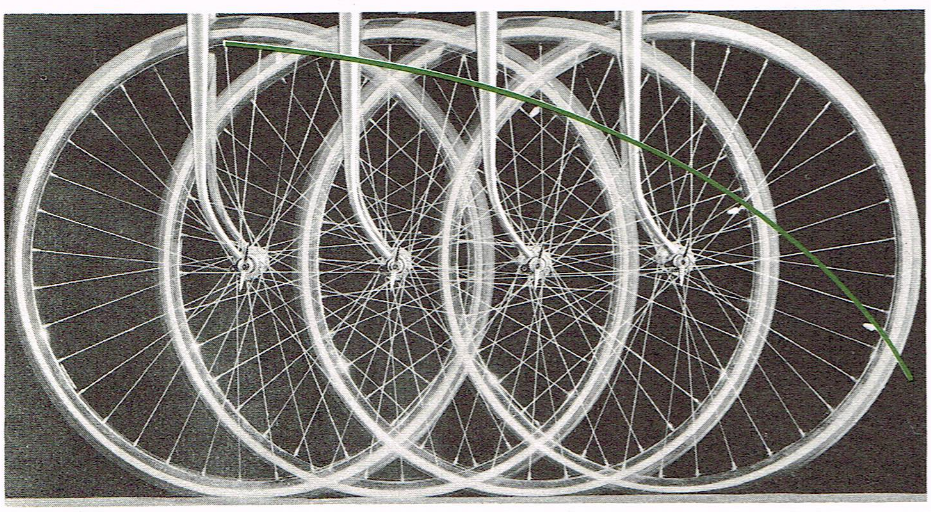


Exemple d'un même mouvement observé par rapport à deux référentiels différents : les habitants de la maison observent le mouvement du parachutiste par rapport au référentiel constitué par la maison ou la Terre; l'équipage de l'avion observe le mouvement du même parachutiste par rapport au référentiel constitué par l'avion.

et si l'on change de référentiel ?

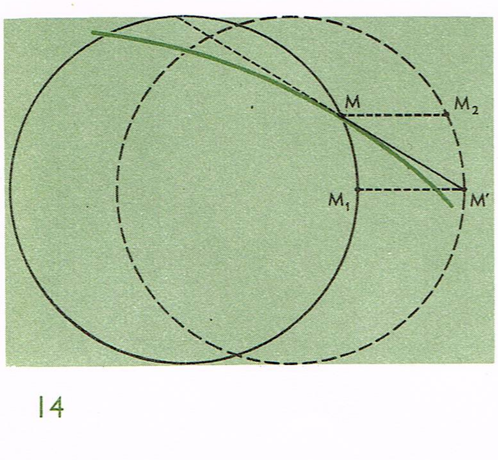
La notion de mouvement étant essentiellement relative au référentiel considéré, le problème fondamental de la cinématique va être le suivant : dans un changement de référentiel, que vont devenir le repérage d'un même mobile, sa vitesse et son accélération ?

Pour ce qui est du temps, qui sert à indiquer l'instant des repérages et qui intervient dans la détermination des vitesses et des accélérations, on adopte en mécanique classique une attitude très simple. Dans ses déplacements, l'observateur continue à faire confiance aux astronomes. Il se préoccupe peu de la façon dont lui parviennent les signaux horaires. Il admet que l'heure qu'il emporte sur sa montre réglée au départ coïncide toujours avec l'heure construite au point de départ. Le temps a pour tous les observateurs une valeur universelle; il est le même pour tous les référentiels, quels que soient leurs mouvements les uns par rapport aux autres.



Chronophotographie d'une roue de bicyclette munie de sa valve. Les différentes images sont prises à des intervalles de temps réguliers.

Le repérage de la position du mobile par rapport à deux référentiels différents est un simple problème de géométrie sur lequel nous n'insisterons pas. Qu'il nous suffise de concevoir clairement que si l'on connaît, à un instant donné, les trois coordonnées x, y, z du parachutiste de la figure par rapport au référentiel R, et, d'autre part, la position du référentiel R′ par rapport à R, il est possible de calculer, ou de faire calculer par une machine, les trois coordonnées x', y', z' du parachutiste, à ce même instant, par rapport au référentiel R′.

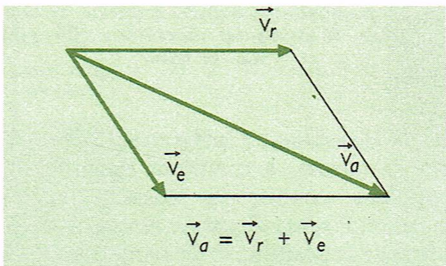


Pendant le petit intervalle de temps Δt qui sépare ces deux positions de la roue, le déplacement de la valve par rapport au cadre de la bicyclette est MM_1 et son déplacement par rapport au sol est MM'. M_2 est la position qu'aurait la valve après le temps Δt si la roue avait été bloquée autour de son axe. On constate que MM' apparaît comme la diagonale du parallélogramme construit sur MM_1 et MM_2. Or, après division par Δt et passage à la limite, $\overrightarrow{MM'}$ fournit la vitesse absolue $\overrightarrow{V_a}$, $\overrightarrow{MM_1}$ la vitesse relative $\overrightarrow{V_r}$ et $\overrightarrow{MM_2}$ ce qu'on appelle la vitesse d'entraînement $\overrightarrow{V_e}$.

La question qui se pose ensuite est la suivante : les vitesses du mobile par rapport à deux référentiels différents n'étant pas les mêmes, comment passer

de l'une à l'autre ? Précisons les données de ce problème, que l'on appelle le problème de **composition des vitesses.** De nos deux référentiels, nous en distinguerons un premier R, que nous appellerons conventionnellement fixe ou absolu, et un second R′, mobile de mouvement relatif à R connu. Quel est alors le lien entre la vitesse du mobile par rapport à R (vitesse absolue $\vec{V}_a$) et la vitesse de ce mobile par rapport à R′ (vitesse relative $\vec{V}_r$) ?

Le lien trouvé sur l'exemple précédent entre la vitesse absolue et la vitesse relative est général. La vitesse absolue s'obtient en composant suivant la règle du parallélogramme la vitesse relative



et la vitesse d'entraînement. On dit que $\vec{V}_a$ est la somme géometrique des deux vecteurs $\vec{V}_r$ et $\vec{V}_e$.

La vitesse d'entraînement est, d'une façon générale, la vitesse absolue qu'aurait le mobile s'il lui était arrivé d'être figé dans le référentiel mobile.

comment se composent les accélérations ?

La réponse à cette question, qui consiste à trouver le lien qui existe entre l'accélération absolue et l'accélération relative, est plus riche de conséquences, mais aussi de difficultés, que celle qui concerne les vitesses. C'est, en effet, aux accélérations que seront reliées, comme nous le verrons, les forces, causes du mouvement. Mais l'accélération d'un mobile est une notion relative, et les forces qui agissent sur lui n'apparaissent pas du tout *a priori* comme une notion relative : que j'observe un satellite artificiel à la lunette d'un observatoire ou à bord d'un avion effectuant vrilles, tonneaux et loopings, il est bien évident que ma présence et les conditions d'observation n'exercent aucune influence sur les forces qui agissent sur ce satellite. Pourtant, les mouvements observés dans ces deux conditions sont bien différents. Comment faudra-t-il donc relier de façon cohérente les accélérations du satellite mesurées dans ces deux conditions d'observation aux forces qui agissent sur lui ? Tel est le genre de question auquel la loi de **composition des accélérations** permet de répondre.

Et si l'on objecte qu'il est ridicule de vouloir « faire de la mécanique » dans un avion effectuant des loopings, je répondrai que c'est malheureusement dans des conditions analogues que nous nous trouvons sur la Terre. Nous sommes tous entraînés avec elle dans une gigantesque farandole : dans son mouvement de rotation autour de sa ligne des pôles, dans son mouvement de rotation autour du Soleil, dans le mouvement du système solaire dans la Galaxie, etc.

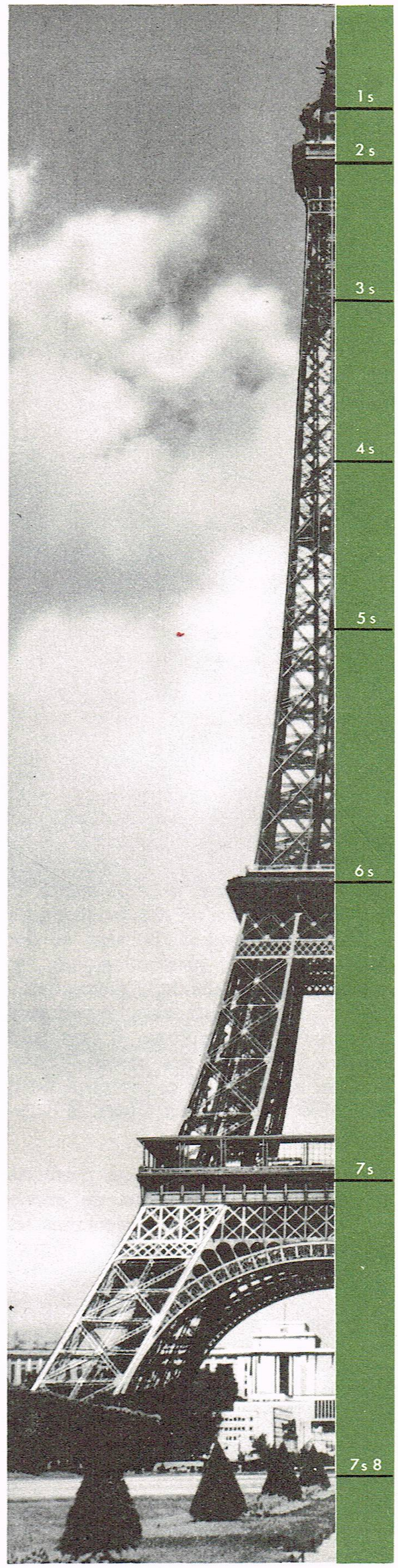
C'est ainsi que l'homme, attaché à la Terre, est aussi mal placé pour comprendre le mouvement d'une planète que l'occupant d'un avion en looping pour comprendre le mouvement d'un train lancé sur une portion rectiligne de rail.

Pour résoudre le problème de la composition des accélérations, il nous faudra donc réfléchir davantage sur les mouvements, en observer attentivement certaines catégories, et cela par rapport à des référentiels différents. C'est ce que nous allons commencer à faire à partir d'observations familières.

le mouvement des graves

Un des mouvements le plus couramment observés est la chute d'un objet à la surface de la Terre. Lâché sans vitesse, l'objet tombe suivant la verticale (par définition même de la verticale). L'étude des positions, puis des vitesses et des accélérations peut être faite en utilisant la chronophotographie, ou un matériel expérimental approprié.

Ces expériences, sur lesquelles nous n'insisterons pas ici, ont un caractère tout à fait fondamental. C'est à partir d'elles que s'est élaborée (Galilée, 1564-1642) la mécanique classique telle que nous la concevons actuellement. Ce sont elles qui demeurent encore à la base des exposés usuels.

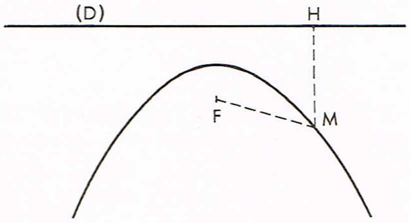


Distances parcourues seconde par seconde par un corps tombant en chute libre de la tour Eiffel.

Phot. Edit. Chantal.



Les jets d'eau sortant d'embouchures fines et inclinées matérialisent les trajectoires paraboliques.



La *parabole* est définie à partir de son foyer F et de sa directrice (D) par MF = MH.

Si la vitesse initiale communiquée au corps n'est ni nulle ni verticale, par exemple dans le lancement d'une pierre, la trajectoire est une parabole.

Ayant fait cette constatation primitive qu'un objet lancé en l'air finit par retomber sur le sol, l'application essentielle que lui a trouvée l'homme a été malheureusement d'essayer de faire retomber cet objet sur la tête de son voisin. Le problème du tir d'une pièce d'artillerie sur un but déterminé est compliqué par le fait qu'à cause de la résistance de l'air la trajectoire n'est plus parabolique; il faut tenir compte du vent et des autres conditions météorologiques, de la mobilité éventuelle du but, etc.; beaucoup d'impondérables interviennent dans les données du problème, il faut faire appel à la théorie des probabilités et tirer par salves. Aussi ce problème a-t-il fait l'objet d'une science à part, la balistique extérieure.

Mais nous avons assisté, ces dernières années, à des mouvements de graves — c'est ainsi que nous désignerons les corps pesants — à des altitudes de plus en plus élevées, où la résistance de l'air se fait de moins en moins sentir. Le vol d'un V2, le vol balistique d'un X-15 qui, coupant ses réacteurs à un certain instant, se met à tomber comme une pierre, présentent des portions de trajectoires assez rigoureusement paraboliques.

Nous reviendrons sur la résistance de l'air. Pour l'instant, n'en tenons pas compte et considérons la catégorie des mouvements rectilignes ou paraboliques des graves dans une région pas trop étendue, située au voisinage de la surface terrestre. De la considération de ces mouvements, nous retiendrons dès à présent le fait essentiel suivant, qui n'a pas toujours paru évident : la position initiale et le vecteur vitesse à l'instant initial déterminent le mouvement ultérieur du projectile. C'est le **principe des conditions initiales,** qui sera valable pour toute catégorie de mouvements.

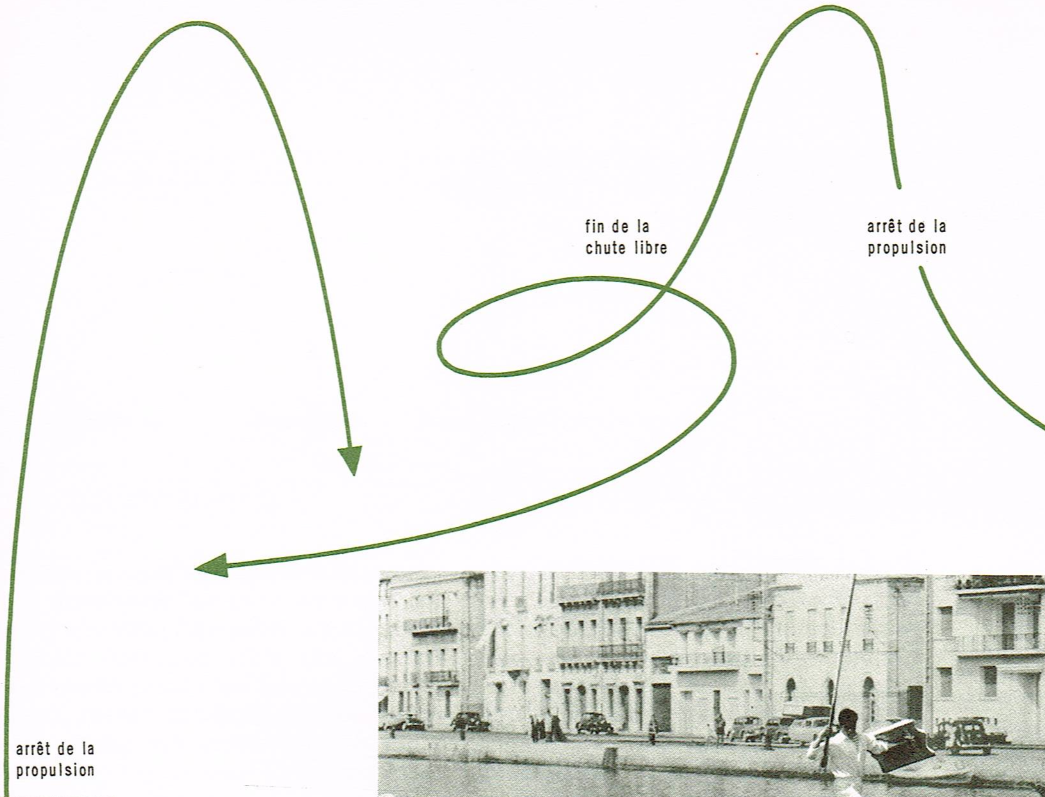
Ces mouvements des graves, rapidement décrits ci-dessus, ne peuvent être considérés que par rapport à un référentiel bien déterminé, qui n'a pas été précisé, mais qui est évidemment constitué par la Terre elle-même.

L'étude des mouvements des graves dans une région déterminée montre que le vecteur vitesse varie d'une façon assez complexe, mais que le vecteur accélération $\vec{g}$ reste constant en direction et sens (verticale descendante) et en grandeur (9,81 m/s^{-2} à Paris), et cela indépendamment de la taille et de la forme du corps d'épreuve. Il est alors naturel de tenter d'expliquer le mouvement par son accélération et de voir ce que donne l'hypothèse suivante : dans une région pas trop étendue à la surface de la Terre, il existe un champ d'accélération $\vec{g}$, qui est constant et que l'on appelle le **champ d'accélération de la pesanteur ;** tout corps situé dans cette région possède automatiquement l'accélération $\vec{g}$. Déterminer les conséquences de cette hypothèse est le travail du mathématicien. Connaissant l'accélération qui est la dérivée de la vitesse, il sait, par l'opération inverse de la dérivation, l'intégration, déterminer la vitesse, à condition toutefois de supposer connue la vitesse initiale. Ensuite, il saura de même remonter de la vitesse à la position, et déterminer celle-ci en fonction du temps, en supposant cette fois connue la position initiale. Le résultat de ces calculs est que, inversement, tout mouvement à accélération constante $\vec{g}$ rentre dans la catégorie des mouvements des graves précédemment décrits.

Remarquons bien que l'hypothèse qui a été faite, suivant laquelle le mouvement est expliqué par l'accélération, est conforme au principe des conditions initiales que nous avons énoncé : l'intégration mathématique du mouvement à partir de son accélération ne fournit une solution bien déterminée que si les conditions initiales de position et de vitesse sont données.

Nous avons donc une explication correcte des mouvements des graves. Il ne faut pas attribuer à cette expression « explication correcte » un sens absolu. Ce que nous voulons dire par là, c'est que nous sommes en mesure de mettre le problème en équations, de le résoudre sur le papier, c'est-à-dire de prévoir les mouvements. Nous nous sommes dégagés du domaine purement empirique, et c'est là une étape essentielle dans la constitution d'une science. Mais pourquoi les corps sont soumis à cette accélération de la pesanteur, nous n'en savons encore rien; nous tenons une explication, au moins partielle, qui nous est commode, et nous n'en demandons pas plus pour l'instant.

Phot. Pougnet (Rapho).

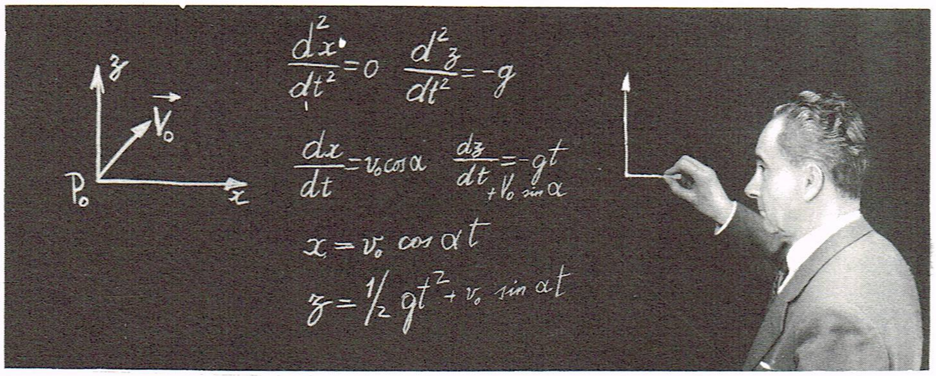


Alors que les trajectoires théoriquement paraboliques d'une pierre ou d'un projectile d'artillerie sont affectées par la présence de l'air, son mouvement et autres impondérables, celles d'une fusée V2 *(à gauche)* et d'un avion Bell X-15 *(ci-dessous)* présentent des portions que l'on peut considérer comme de vrais arcs de parabole.

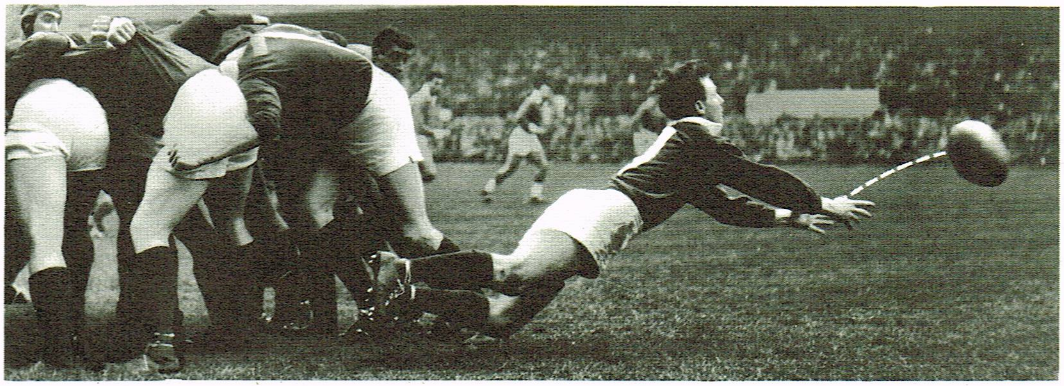




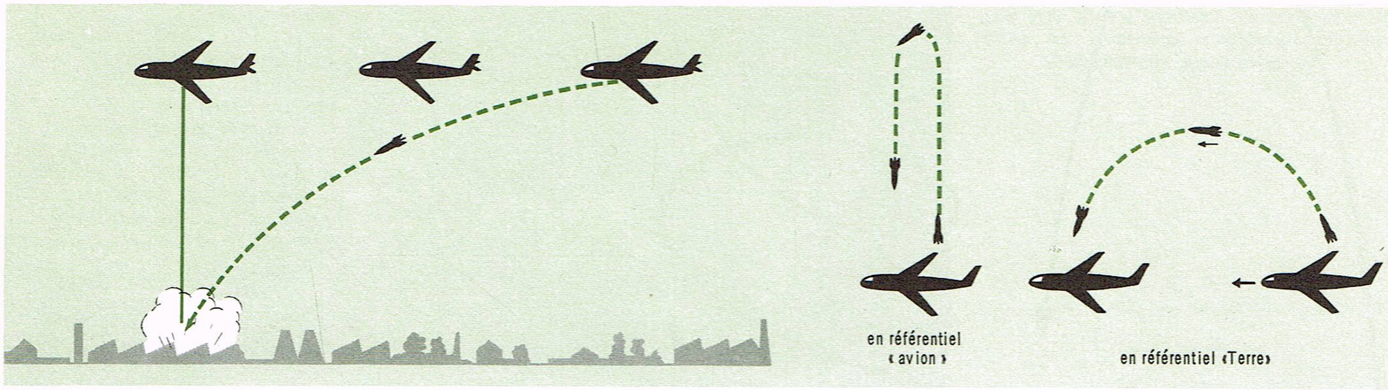




L'expérimentateur a décelé le champ constant de l'accélération de la pesanteur; le mathématicien calcule les mouvements qui ont lieu dans ce champ; il trouve des mouvements que l'expérimentateur sait réaliser.



Phot. U. S. I. S., Jacques Lang (Rapho), Larousse, Presse-Sports.



un premier changement de référentiel

Un avion en vol horizontal rectiligne uniforme lâche une bombe à un instant déterminé. La vitesse initiale de la bombe, par rapport à la Terre, est un vecteur horizontal équipollent (égal en grandeur, direction et sens) à la vitesse de l'avion. La trajectoire de la bombe par rapport à la Terre est donc une parabole, et celui qui voit la bombe se détacher lorsque l'avion passe à la verticale du lieu où il se trouve ne risque pas d'être atteint par elle. Par rapport au référentiel constitué par l'avion, la vitesse initiale de la bombe est nulle et sa trajectoire est la verticale attachée à l'avion. L'équipage voit constamment la bombe au-dessous de l'avion et l'impact se produit au point survolé par l'avion lorsque la bombe atteint le sol. Les études du mouvement de la bombe faites au sol par les observateurs terrestres, et à bord de l'avion par l'équipage, montrent ainsi que l'accélération par rapport à la Terre et l'accélération par rapport à l'avion sont les mêmes.

L'explication du mouvement par l'accélération de la pesanteur est la même dans le référentiel « Terre » et dans le référentiel « avion ». Si, au lieu de lâcher une bombe sans vitesse initiale, on tire à partir de l'avion un obus suivant sa verticale ascendante, l'obus montera, puis redescendra sur la verticale de l'avion et retombera sur lui (mécanique du mouvement des graves dans le référentiel « avion»). Pour un observateur terrestre, la trajectoire de l'obus sera parabolique, son trajet horizontal étant le même que celui de l'avion lorsqu'il se retrouvera à l'altitude de l'avion (mécanique du mouvement des graves dans le référentiel « Terre »).

En généralisant le résultat obtenu dans cet exemple pour l'appliquer à un mouvement quelconque, nous obtenons une première réponse à notre problème de composition des accélérations : si le mouvement du référentiel mobile R′ par rapport au référentiel absolu R est une translation rectiligne uniforme, l'accélération absolue est égale à l'accélération relative.

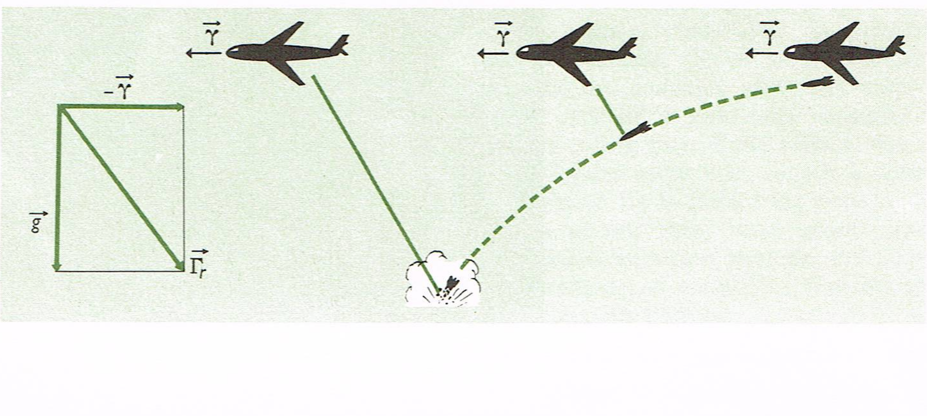
un second changement de référentiel

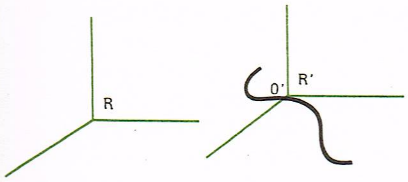
Imaginons maintenant que le mouvement de l'avion, toujours horizontal et rectiligne, n'est plus uniforme, mais uniformément accéléré, c'est-à-dire qu'il a, par rapport à la Terre, une accélération $\vec{\gamma}$ constante. Le mouvement de la bombe par rapport à la Terre est le même mouvement parabolique que précédemment (en supposant que la vitesse de l'avion au moment du largage soit la même que précédemment). Mais l'équipage de l'avion observe cette fois que la chute s'opère selon une droite inclinée et que l'accélération $\vec{\Gamma}_r$ s'obtient en faisant la somme géométrique du vecteur $\vec{g}$ (accélération absolue) et du vecteur $-\vec{\gamma}$, opposé à $\vec{\gamma}$. Appelons maintenant « accélération d'entraînement $\vec{\Gamma}_e$ » l'accélération (absolue) $\vec{\gamma}$ de l'avion. Nous aurons dans ce cas une loi de composition des accélérations qui possède le même caractère de simplicité que la loi générale de composition des vitesses : l'accélération absolue est la somme géométrique de l'accélération relative et de l'accélération d'entraînement. Au lieu de $\vec{\Gamma}_r = \vec{g} - \vec{\gamma}$, nous pouvons en effet écrire $\vec{\Gamma}_a = \vec{\Gamma}_r + \vec{\Gamma}_e$.

Cet énoncé, obtenu sur un exemple simple, se généralise sans modification pour un mouvement quelconque, à condition toutefois que le mouvement du référentiel mobile R′ par rapport au référentiel absolu R soit un mouvement de translation.

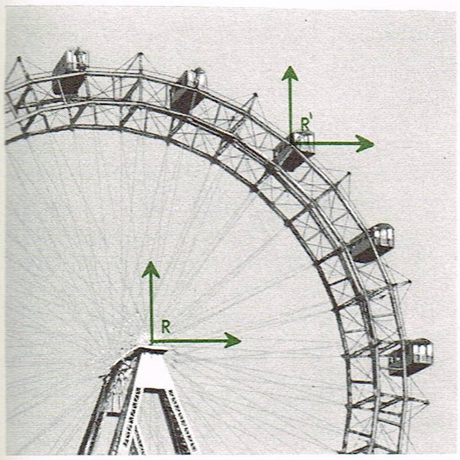
L'accélération d'entraînement, que nous venons de définir dans le seul cas où le mouvement de R′ par rapport à R est un mouvement de translation général, existera toujours. Dans un mouvement absolument quelconque de R′ par rapport à R, la définition en est analogue à celle de la vitesse d'entraînement : l'**accélération d'entraînement** est l'accélération absolue qu'aurait le mobile s'il lui était arrivé d'être figé dans le référentiel mobile.

L'étude de l'accélération d'entraînement étant ainsi terminée, nous n'en avons pas fini pour autant avec la loi de composition des accélérations. Dans les cas

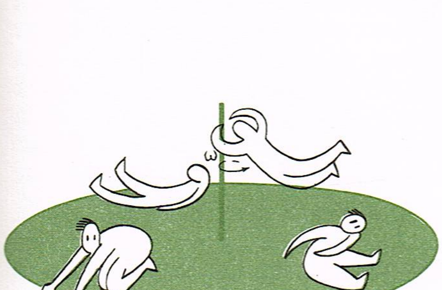




Un mouvement de translation général de R′ par rapport à R est un mouvement dans lequel les droites de R′ gardent les directions fixes par rapport à R. Le point O′ de R′ a, par rapport à R, un mouvement absolument quelconque, mais tous les points de R′ ont, à un instant donné, le même vecteur vitesse et le même vecteur accélération, qui est l'accélération d'entraînement.

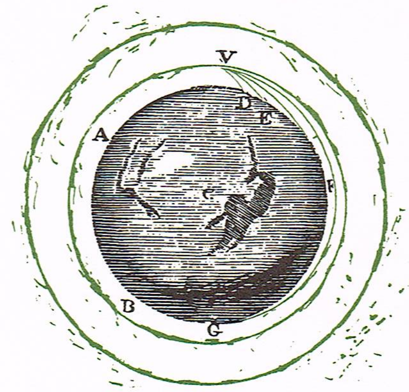


Une des cabines de la grande roue a un mouvement de translation. Tous ses points ont la même vitesse que le point d'attache dont le mouvement est circulaire uniforme. Tous ses points ont la même accélération que le point d'attache, laquelle est centripète et de grandeur $\omega^2 r$ (ω vitesse angulaire de rotation de la roue, r rayon de la roue).



Accélération d'entraînement dans le cas ou le mouvement de R′ par rapport à R est un mouvement de rotation autour d'un axe fixe. Ici, elle est variable, à un instant donné, avec la place occupée par le mobile dans R′.

où le mouvement de R′ par rapport à R n'est pas un mouvement de translation, cette loi n'aura plus le caractère de simplicité que nous lui avons trouvé jusqu'ici. L'accélération absolue ne sera plus la somme géométrique de l'accélération relative et de l'accélération d'entraînement. Il faudra encore ajouter un terme complémentaire : l'accélération de Coriolis (1792-1843). Pour la définir, les mouvements des graves et les mouvements rectilignes d'un avion ne vont plus nous suffire. Nous allons d'abord faire appel à une autre catégorie de mouvements, très intéressants à beaucoup de points de vue : ceux des satellites artificiels de la Terre.



Newton déjà avait prévu la satellisation (gravure extraite des *Principia*).

le mouvement des satellites artificiels de la Terre

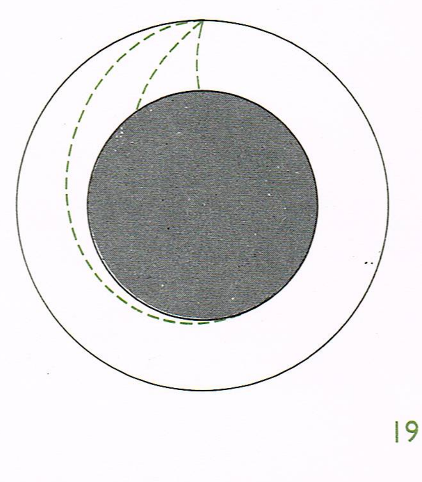
Il est évident que l'on augmente la portée d'une pièce d'artillerie en augmentant la vitesse initiale de l'obus lancé. La question se pose alors de savoir ce qui se passerait si l'on pouvait augmenter indéfiniment cette vitesse initiale. Newton lui-même s'était posé la question et avait prévu que, pour une vitesse initiale assez grande, le projectile ne retomberait plus sur la terre.

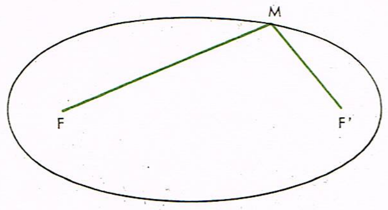
Les expériences de ces dernières années ont confirmé les idées de Newton, et nous pouvons à présent décrire des expériences de tir qui conduisent à l'envoi dans l'espace d'un engin spatial, sans plus faire appel à l'imagination d'un Jules Verne, mais en prenant pour base des faits réels précis. Pour notre description, nous nous placerons dans les conditions particulières suivantes : nous supposerons que la position de tir de l'engin sera située à 200 km d'altitude (on est alors pratiquement sorti de l'atmosphère terrestre) et que la vitesse initiale communiquée à l'engin sera horizontale, sans nous préoccuper pour l'instant ni de la méthode employée pour faire parvenir l'engin à cette position, ni de la façon dont on lui communiquera sa vitesse initiale. Décrivons rapidement les mouvements possibles de l'engin.

Pour une vitesse initiale nulle ou assez faible, l'engin retombe rapidement sur la Terre au voisinage du point au-dessus duquel se trouve la position de lancement. Puis, pour des vitesses de plus en plus grandes, l'engin retombe en un point de plus en plus éloigné. Il pourra dépasser le point situé aux antipodes et commencer à s'enrouler autour de la Terre. Nous ne nous intéresserons pas plus longuement à ces trajectoires pour lesquelles les engins retombent dans l'atmosphère, en y subissant fortement la résistance de l'air, et finalement atteignent le sol.

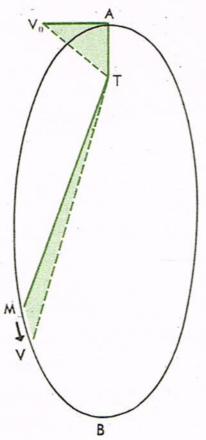
A partir d'une certaine vitesse, le cap est franchi et nous obtenons une trajectoire circulaire, sur laquelle on peut observer que le mouvement est uniforme : l'engin est satellisé. La vitesse initiale correspondant à ce cas est appelée la **première vitesse spatiale.** Elle s'élève (pour une altitude de 200 km) à 7,8 km/s. Le mouvement est périodique. La période est donnée, en fonction de la vitesse v et du rayon r de la trajectoire, par la formule $T = \frac{2\pi r}{v}$; elle est donc ici d'environ 88 minutes. L'accélération est centripète et a pour grandeur, constante, $\frac{v^2}{r}$.

Pour des vitesses initiales dépassant la première vitesse spatiale, la trajectoire s'allonge. Elle devient elliptique et le mouvement est appelé « képlérien ». Il

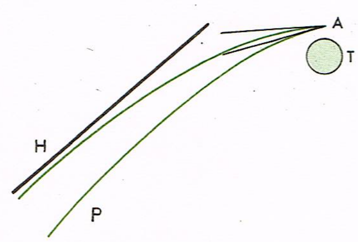




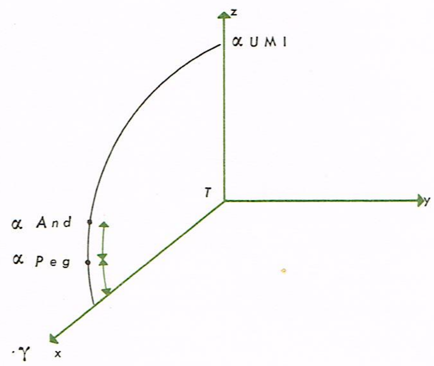
Ci-dessus, l'ellipse est définie à partir de ses deux foyers F et F' par $MF + MF' = C^{te}$



Le vitesse du satellite décroît du point A, le plus proche de la Terre (périgée), au point B, le plus éloigné (apogée).



Trajectoire parabolique (P) et trajectoire hyperbolique (H), qui se rapproche d'une certaine droite appelée *asymptote*.



Le référentiel « Terre-Étoiles » Txyz est un trièdre trirectangle; T est le centre de la Terre, Tz est dirigée vers l'étoile polaire, Tx vers le point vernal.

satisfait, en effet, aux deux premières lois que Kepler (1571-1630) a énoncées en 1609 pour le mouvement d'une planète autour du Soleil. La première affirme que la trajectoire est une ellipse dont un des foyers est le centre de la Terre. La seconde indique la façon dont la trajectoire est décrite en fonction du temps;
c'est suivant la **loi des aires :** le rayon vecteur TM balaie des aires égales en des temps égaux. Le mouvement est encore périodique.

Si, toujours dans les mêmes conditions de départ, on continue à augmenter la vitesse initiale, l'ellipse s'allonge de plus en plus et l'engin atteint des points de plus en plus éloignés de la Terre. Alors un second cap se franchit : pour une certaine vitesse, le mouvement cesse d'être périodique, l'engin s'éloigne indéfiniment de la Terre, la trajectoire devient parabolique, l'engin échappe à la Terre; il est véritablement injecté dans l'espace. La vitesse correspondant à ce cas est appelée la **seconde vitesse spatiale.** Elle est (à nos 200 km d'altitude) de 11 km/s. Pour des vitesses initiales encore plus grandes, l'engin continue à s'éloigner indéfiniment de la Terre sur une trajectoire qui devient hyperbolique. Si v_1 et v_2 sont respectivement la première et la seconde vitesse spatiale, toute vitesse comprise entre v_1 et v_2 est une **vitesse de satellisation,** toute vitesse supérieure à v_2, une **vitesse d'évasion.**

Naturellement, dans la réalité, les choses ne sont pas aussi simples, et nous n'avons donné qu'une description schématique de ce qui se passerait si la Terre était isolée dans l'espace et si elle était rigoureusement sphérique et homogène. En réalité, la présence de la Lune, du Soleil, etc., va modifier petit à petit le mouvement elliptique képlérien du satellite terrestre; la trajectoire parabolique ou hyperbolique de l'engin spatial ne sera suivie que dans le voisinage de la Terre. Le fait que la Terre n'est pas rigoureusement sphérique et que les masses n'y sont pas réparties de façon homogène perturbe également les mouvements. Mais nous ne ferons intervenir ces faits que plus tard, et c'est sur cette description schématique que nous voulons d'abord raisonner.

quel en est le référentiel ?

Ici encore, il y a quelque chose que nous n'avons pas précisé, alors que nous avons insisté sur la nécessité de le faire : c'est le référentiel utilisé pour la description des mouvements précédents. Contrairement à ce qui se passait lors de notre description des mouvements des graves, l'observateur terrestre est à présent mal placé en face de ces mouvements spatiaux. Il nous faut faire appel à un référentiel dans lequel la Terre jouera encore un rôle essentiel, mais par rapport auquel la rotation de la Terre dans l'espace sera mise en évidence. C'est aux étoiles qu'il convient de s'adresser, et le référentiel convenable sera constitué par le centre de la Terre et par la figure géométrique invariable formée par les étoiles sur la sphère céleste. Ce référentiel sera fréquemment utilisé. Nous appellerons donc **référentiel Terre-Étoiles** le référentiel matérialisé par un trièdre ayant son origine au centre de la Terre et ses axes dirigés vers des étoiles fixes.

C'est par rapport à ce référentiel Terre-Étoiles que les mouvements des satellites artificiels de la Terre et des engins spatiaux satisfont à la description schématique précédente.

Si l'on se borne à considérer la trajectoire circulaire, elle s'effectue dans un plan passant par le centre de la Terre et par la position de lancement, et contenant le vecteur vitesse initial (qui était un vecteur vitesse par rapport au référentiel Terre-Étoiles). Ce plan demeure fixe par rapport au référentiel Terre-Étoiles. Pour les observateurs terrestres, les choses sont beaucoup moins simples : la Terre tourne, en les entraînant, et la trajectoire du satellite leur paraît tourner dans un sens opposé; ils voient donc la trajectoire du satellite s'enrouler autour de la Terre à la façon d'une pelote de ficelle, dans un mouvement complexe.

quelle peut en être l'explication ?

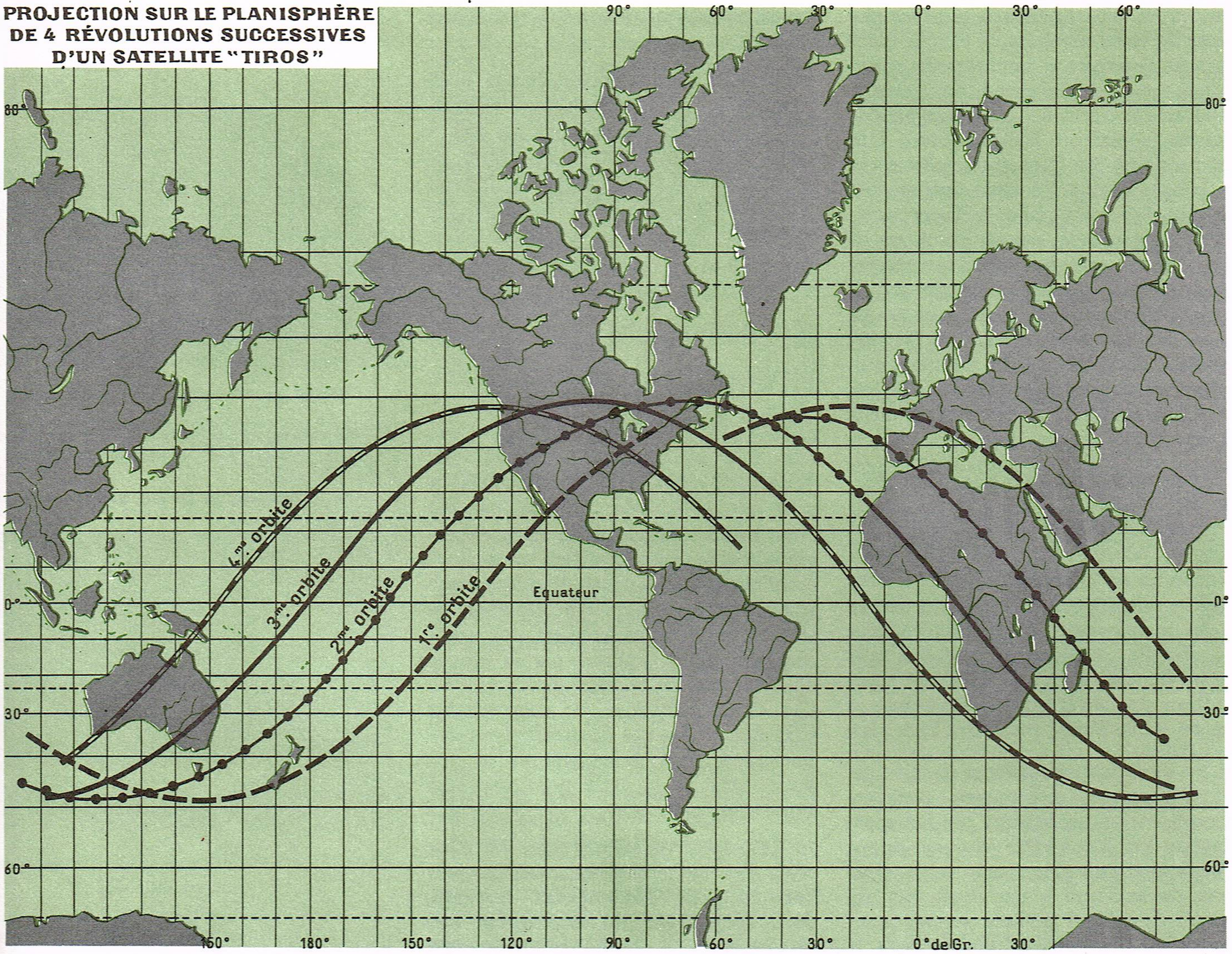
Essayons à présent d'obtenir une certaine explication de ces mouvements des satellites terrestres par rapport au référentiel Terre-Étoiles.

Il nous faut pour cela poursuivre l'étude



Ci-contre, photographie mettant en évidence la rotation du référentiel « Terre-Étoiles » par rapport au référentiel « Terre ».

Phot. Larousse.



de ces mouvements en supposant que nous avons à notre disposition plusieurs engins, d'importance variée, que nous pouvons satelliser à des altitudes variées. Mais il nous suffira de considérer les trajectoires circulaires, c'est-à-dire d'indiquer des résultats relatifs à la première vitesse spatiale.

Le premier fait essentiel constaté est que la première vitesse spatiale ne dépend pas de l'engin, mais uniquement de l'altitude du point de lancement. A une altitude donnée, la vitesse nécessaire à la satellisation suivant une trajectoire circulaire ne dépend ni de la taille, ni de la forme, ni des matériaux de construction de l'engin. Mais plus l'altitude du point de lancement est élevée, plus facile est la satellisation. Le tableau ci-contre donne les premières vitesses spatiales V_1 correspondant à diverses altitudes h.

Il y est également indiqué la distance r du satellite au centre de la Terre, c'est-à-dire le rayon de son orbite circulaire, en prenant pour rayon de la Terre sa valeur moyenne (6 370 km). On peut alors trouver la loi de variation de la première vitesse spatiale. Le produit $V_1^2 r$ garde toujours la même valeur. La première vitesse spatiale est inversement proportionnelle à la racine carrée de la distance du satellite au centre de la Terre : $V_1^2 = \frac{\mu}{r}$.

Le tableau ci-dessous indique également un élément important du mouvement circulaire uniforme : sa période. La période T est donnée par la formule $T = \frac{2\pi r}{v}$, avec $v = V_1 = \sqrt{\frac{\mu}{r}}$. On constate que le rapport $\frac{T^2}{r^3} = \frac{4\pi^2}{\mu}$ est le même pour toutes les trajectoires circulaires. Les carrés des temps de révolution sont proportionnels aux cubes des rayons des orbites. Cela correspond, dans le cas particulier des trajectoires circulaires, à la troisième loi que Kepler a énoncée en 1619, dix ans après avoir

h (km)	r (km)	V_1 (km/s)	$V_1^2 r = \mu$	T (période en minutes)
200	6 570	7,78	40×10^6	88,2
400	6 770	7,67	40×10^6	93,2
600	6 970	7,56	40×10^6	96,4
800	7 170	7,45	40×10^6	100,6
1 000	7 370	7,35	40×10^6	104,8

énoncé les deux premières, pour les planètes du système solaire. Si chaque mouvement circulaire a une période déterminée, on peut chercher inversement le mouvement circulaire ayant une période donnée (supérieure à 88 minutes). Une des périodes les plus intéressantes est celle de 24 heures. Un engin ainsi satellisé dans le plan équatorial de la Terre survolerait toujours le même point de la Terre (avec toujours la même réserve concernant le caractère schématique de notre description). Il serait, avec un ou deux de ses congénères, un élément essentiel du problème des télécommunications spatiales. Ce satellite de 24 heures aura une orbite de 42 000 km de rayon et une vitesse de 3,4 km/s.

En vue de l'explication que nous cherchons, étudions la vitesse et l'accélération du satellite sur sa trajectoire circulaire. Vitesse et accélération sont constantes en grandeur; mais le vecteur vitesse, tangent à la trajectoire, prend au cours du mouvement toutes les directions du plan de l'orbite et ne peut être relié à rien de fixe, alors que le vecteur accélération est constamment dirigé vers le centre de la Terre, point fixe bien réel.

Il est alors naturel, comme dans le cas du mouvement des graves, d'essayer d'expliquer le mouvement par son accélération et d'examiner ce que donne l'hypothèse suivante : tout corps situé hors de la Terre à une distance r du centre de celle-ci possède automatiquement une accélération dirigée vers le centre de la Terre et de grandeur $\frac{\mu}{r^2}$. Nous dirons qu'un champ d'accélération entoure la Terre, et nous l'appellerons le **champ d'accélération de la gravitation terrestre.** Ici encore, ce sera le travail du mathématicien de déterminer les conséquences de cette hypothèse, qui est conforme au principe des conditions initiales. Il remontera, par intégration, et en utilisant les conditions initiales de position et de vitesse, de l'accélération à la vitesse, puis de la vitesse à la position. Le résultat des calculs est que, inversement, tout mouvement à accélération centripète $\frac{\mu}{r^2}$ entre dans la catégorie des mouvements précédemment décrits : mouvements circulaires uniformes si la vitesse initiale est la première vitesse spatiale de l'altitude du point de départ et si elle est perpendiculaire à la droite joignant le centre de la Terre à ce point de départ, mouvements képlériens elliptiques, paraboliques ou hyperboliques si les conditions initiales sont autres.

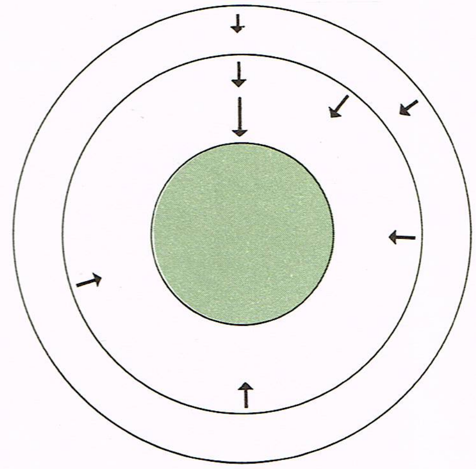
Nous avons donc ici encore, comme pour le mouvement des graves, une explication correcte des mouvements des engins spatiaux au voisinage de la Terre, en ce sens qu'elle nous permet de prévoir le mouvement d'un engin lancé dans des conditions données.

le mouvement des graves et le mouvement des satellites terrestres sont-ils de même nature ?

Nous avons considéré deux catégories de mouvements, tous issus d'un lancement : le mouvement des graves, expliqué localement par le champ d'accélération de la pesanteur; le mouvement des satellites terrestres, expliqué par le champ d'accélération de la gravitation terrestre. Ces deux catégories de mouvements sont-elles de natures essentiellement distinctes, ou peut-on les grouper en une seule catégorie et avoir pour elles une explication commune?

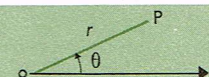
La principale différence entre les deux catégories de mouvements réside dans l'existence de l'atmosphère terrestre, dont nous avons fait abstraction lors de l'étude de la chute des graves; c'est aussi pour pouvoir négliger le freinage dû à l'atmosphère terrestre que nous n'avons envisagé des satellisations qu'à partir de 200 km d'altitude. De même, la nécessité de considérer le champ de pesanteur comme constant nous a obligés à nous cantonner dans une région suffisamment petite pour que l'approximation ainsi faite fût raisonnable.

Pour faire le raccord entre les deux catégories de mouvements, partons des mouvements les plus généraux, ceux des satellites, et abaissons-les par la pensée jusqu'à la surface de la Terre en faisant abstraction de l'atmosphère. On peut concevoir alors qu'il existe une première vitesse spatiale à l'altitude zéro. Si l'on communiquait à un corps situé tout près de la surface de la Terre, supposée privée de son atmosphère, une telle vitesse, il aurait un mouvement circulaire uniforme et raserait continuellement la Terre (un tel satellite rasant peut être envisagé pour la Lune, qui est dénuée d'atmosphère) à la vitesse



Le champ d'accélération de la gravitation terrestre.

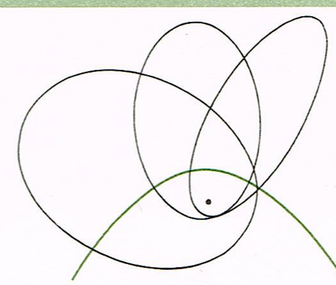
***Ci-dessous,* le calcul des trajectoires.**



$$\vec{\Gamma} = -\frac{\mu}{r^3}\vec{OP}$$

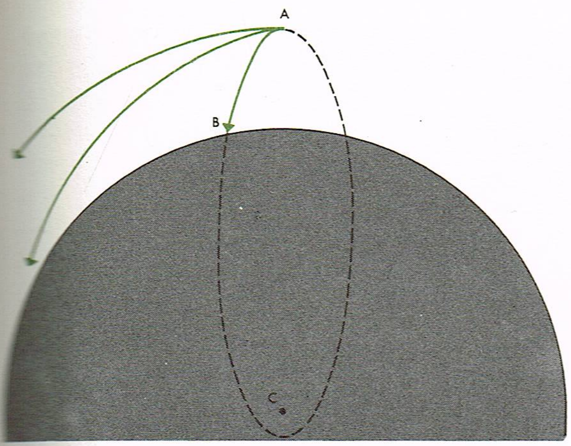
$$\begin{cases} \frac{1}{2}m(r'^2 + r^2\theta'^2) = \frac{\mu}{r} + h \\ r^2\theta' = C \end{cases}$$

$$r = \frac{p}{1 + e\cos(\theta - \alpha)}$$

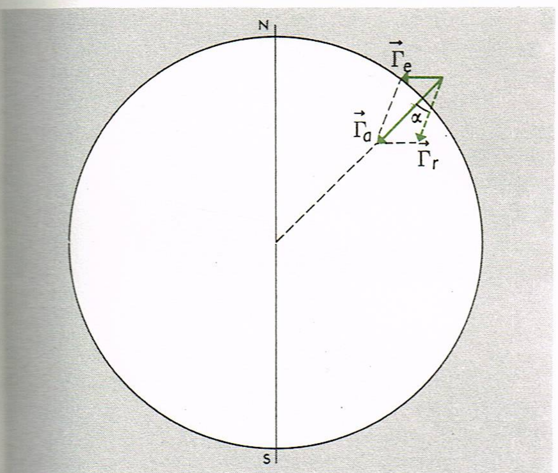


L'observation des trajectoires contrôle parfaitement les calculs.

de 7,92 km/s et avec une période de 84,2 minutes. Mais si l'on communiquait au corps proche de la surface terrestre une vitesse inférieure à cette première vitesse spatiale de l'altitude zéro, il aurait un mouvement elliptique képlérien, vite interrompu par sa rencontre avec la surface de la Terre. Comme nous nous trouvons là dans le domaine du mouvement des graves, nous devons donc considérer les mouvements paraboliques des graves à la surface de la Terre comme un cas particulier approximatif des mouvements elliptiques képlériens. Cela est assez naturel et peut être précisé par le calcul. L'ellipse képlérienne, dont nous ne considérons qu'un petit morceau au voisinage du point de lancement, et dont un foyer, au centre de la Terre,



Comment un satellite devient un grave.



L'accélération de la pesanteur $\vec{\Gamma}_r$ est l'accélération $\vec{\Gamma}_a$ corrigée de $-\vec{\Gamma}_e$.
$\vec{\Gamma}_e$ rencontre perpendiculairement la ligne des pôles et a pour grandeur $\omega^2\rho$ où ω est la vitesse angulaire de rotation de la Terre et ρ le rayon du parallèle terrestre considéré. Nulle aux pôles et maximale à l'équateur, cette correction fait varier l'accélération de la pesanteur de la valeur 9,83 m/sec^2 au pôle à la valeur 9,78 à l'équateur et introduit une légère *déviation α de la verticale* par rapport à ce rayon terrestre.

est très éloigné dans l'expérience considérée, peut être assimilée à un arc de parabole.

Les mouvements des deux catégories que nous venons de fondre en une seule peuvent alors recevoir une explication commune. Il n'y a pour cela qu'à prétendre que l'accélération de la pesanteur n'est autre que l'accélération de la gravitation terrestre aux altitudes voisines de zéro. Cette affirmation se contrôle par le calcul de $\frac{\mu}{R^2}$, où R est le rayon de la Terre (6 370 km); on trouve, avec la valeur de μ qui figure dans le tableau ci-dessus, 9,8 m/s^{-2}, ce qui correspond assez bien aux valeurs expérimentales de l'accélération de la pesanteur. Ainsi la pesanteur est un aspect du phénomène général de la gravitation. C'est à cette notion essentielle, que nous avons essayé de dégager *a priori*, en hommes du XXe siècle environnés d'une série déjà impressionnante de satellites artificiels, que Newton est arrivé en considérant, selon une légende très vraisemblable, la chute d'une pomme d'un pommier de Woolsthorpe et le mouvement du seul satellite terrestre qu'il pouvait considérer, la Lune.

mais les référentiels étaient différents !

Une fois de plus, nous venons de négliger la question primordiale du référentiel. Ce que nous avons fait n'est donc qu'approximatif, et c'est en revenant à la question des référentiels que nous allons enfin pouvoir en terminer avec la délicate question de la composition des accélérations.

Prenons toujours pour point de départ les mouvements les plus généraux, ceux des satellites. Ils ont été décrits dans le référentiel Terre-Étoiles que nous considérons comme référentiel absolu, selon la terminologie que nous avons adoptée pour le problème du changement de référentiel. Les mouvements des graves ont été décrits dans le référentiel Terre, que nous considérons comme référentiel mobile. Le mouvement du référentiel Terre par rapport au référentiel Terre-Étoiles est un mouvement de rotation uniforme autour de la ligne des pôles. Un tour s'effectuant en 24 heures (86 400 secondes), la vitesse angulaire de ce mouvement de rotation vaut, en unités C. G. S., $\frac{2\pi}{86\,400}$, soit moins de 10^{-4}. Les mouvements des graves sont donc expliqués par une accélération relative, alors que les mouvements des satellites sont expliqués par une accélération absolue. Or, nous avons vu précédemment qu'il y avait toujours à tenir compte dans un changement de référentiel d'une accélération d'entraînement. Nous avons vu, d'autre part, quelle était cette accélération d'entraînement dans le cas où le mouvement du référentiel mobile R′ par rapport au référentiel absolu R était un mouvement de rotation autour d'un axe fixe.

Cela nous permet d'obtenir une première amélioration de notre raccord entre mouvements des graves et mouvements des satellites. L'accélération de la pesanteur n'est pas purement et simplement l'accélération de gravitation de grandeur $\Gamma_a = \frac{\mu}{R^2}$. Elle est, en fait, la différence géométrique de cette accélération et de l'accélération d'entraînement ($\vec{\Gamma}_r = \vec{\Gamma}_a - \vec{\Gamma}_e$).

Ainsi la considération de l'accélération d'entraînement nous apporte une précision sur l'accélération de la pesanteur que nous avons considérée comme une donnée d'expérience; c'est l'accélération de la gravitation terrestre corrigée par un terme qui provient de la rotation de la Terre.

l'accélération complémentaire

Nous avons déjà signalé que la loi de composition des accélérations n'avait pas, en dehors du cas où le mouvement du référentiel mobile par rapport au référentiel absolu est un mouvement de translation, la formulation simple $\vec{\Gamma}_a = \vec{\Gamma}_r + \vec{\Gamma}_e$.

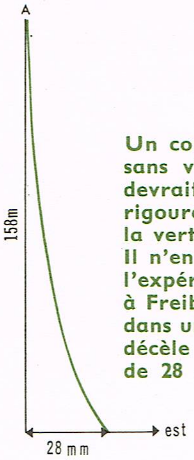
C'est à partir du référentiel absolu Terre-Étoiles et du référentiel mobile Terre que nous allons pouvoir introduire le terme complémentaire qu'il convient de considérer.

En effet, la loi $\vec{\Gamma}_a = \vec{\Gamma}_r + \vec{\Gamma}_e$ appliquée aux référentiels Terre-Étoiles et Terre ne suffit pas pour rendre compte de phénomènes expérimentaux dont les grandeurs assez faibles, mais nettement décelables, ont souvent une importance considérable.

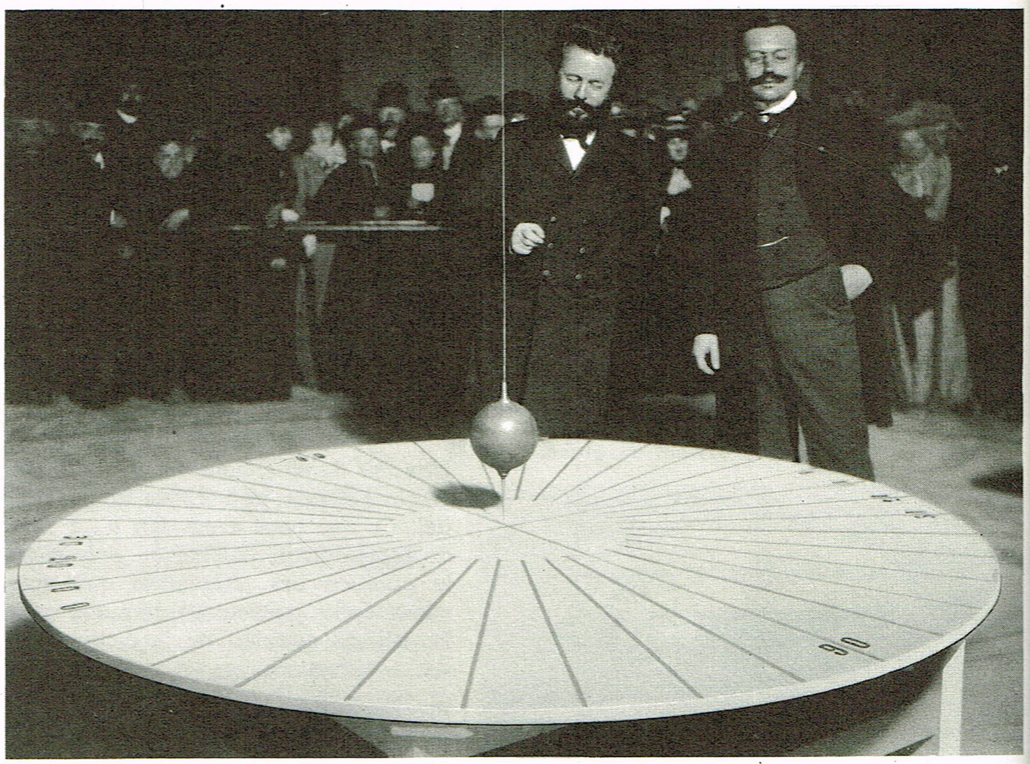
Si cette loi était exacte, l'accélération relative serait contenue dans le plan méridien et tout mouvement pour lequel la vitesse initiale serait contenue dans ce plan aurait lieu rigoureusement dans ce plan. Diverses expériences montrent qu'il n'en est pas ainsi.

Il existe donc une accélération complémentaire, qui rend compte de ces faits expérimentaux. Dans l'expérience de la chute libre, elle est perpendiculaire au plan méridien et dirigée vers l'ouest. D'une façon plus précise et plus générale, elle est perpendiculaire à la fois à l'axe de rotation du référentiel R′ par rapport à R et au vecteur vitesse relative du corps en mouvement, et son sens et sa grandeur peuvent être déterminés comme l'indique la figure en bas de la page suivante.

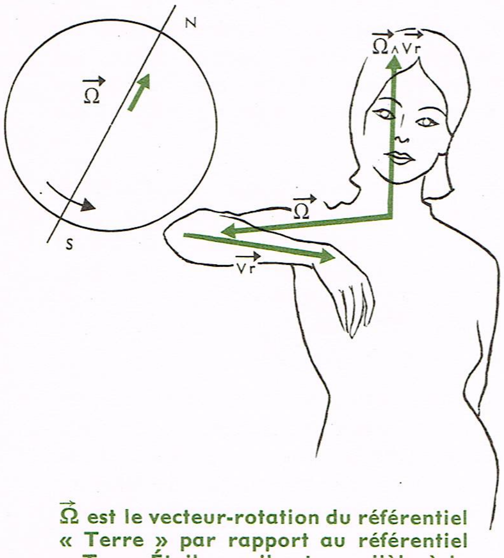
Une telle accélération complémentaire existe dans tous les cas où le mouvement de R′ par rapport à R n'est pas un mouvement de translation, et où l'on peut donc mettre en évidence un vecteur rotation de R′ par rapport à R. Elle se



Un corps lâché sans vitesse initiale devrait décrire rigoureusement une droite : la verticale du point de départ. Il n'en est pas ainsi; l'expérience de Reich, à Freiberg, en 1831, dans un puits de mine de 158 m, décèle une déviation vers l'est de 28 mm.



Ci-contre, reproduction de l'expérience de Foucault réalisée par Flammarion et Berget, en 1902, au Panthéon. La masse pendulaire, lâchée dans le plan méridien, devrait rester dans ce plan. Il n'en est pas ainsi; l'expérience décèle une rotation du plan d'oscillation du pendule.



$\vec{\Omega}$ est le vecteur-rotation du référentiel « Terre » par rapport au référentiel « Terre-Étoiles », il est parallèle à la ligne des pôles, dirigé du pôle Sud vers le pôle Nord, et a pour grandeur $\omega = 2\pi/86400$.

$\vec{V}_r$ est le vecteur vitesse relative du mobile. Le *produit vectoriel* de $\vec{\Omega}$ par $\vec{V}_r$, noté $\vec{\Omega}\Lambda\vec{V}_r$, est un vecteur perpendiculaire à $\vec{\Omega}$ et à $\vec{V}_r$; son sens est tel qu'un « bonhomme d'Ampère » placé dans ce sens voit $\vec{V}_r$ à la gauche de $\vec{\Omega}$; sa grandeur est égale à l'aire du parallélogramme construit sur $\vec{\Omega}$ et $\vec{V}_r$. Alors, l'accélération complémentaire $\vec{\Gamma}_c$ est égale à 2 $\vec{\Omega}\Lambda\vec{V}_r$.

calcule toujours à l'aide de la formule indiquée sous la figure précédente.

Ainsi nous voici enfin au bout de nos peines en ce qui concerne la délicate et fondamentale question de la composition des accélérations : l'accélération absolue est reliée à l'accélération relative par la formule
$\vec{\Gamma}_a = \vec{\Gamma}_r + \vec{\Gamma}_e + \vec{\Gamma}_c$ $(\vec{\Gamma}_c = 2\vec{\Omega}\Lambda\vec{V}_r)$. [1]

Il était indispensable de résoudre de façon précise ce problème, car cette accélération absolue va être reliée de façon simple aux forces, causes du mouvement. Et il est nécessaire que les observateurs qui utilisent une accélération relative sachent de façon précise comment cette accélération est reliée à l'accélération absolue.

l'obus tiré suivant la verticale

Avant d'aborder l'étude des forces, revenons sur l'une des questions que nous nous étions posées au début de cet article, à savoir : est-ce qu'un projectile tiré rigoureusement suivant la verticale va retomber dans la bouche à feu, même compte tenu de la rotation de la Terre. (mais non de la résistance de l'air)?

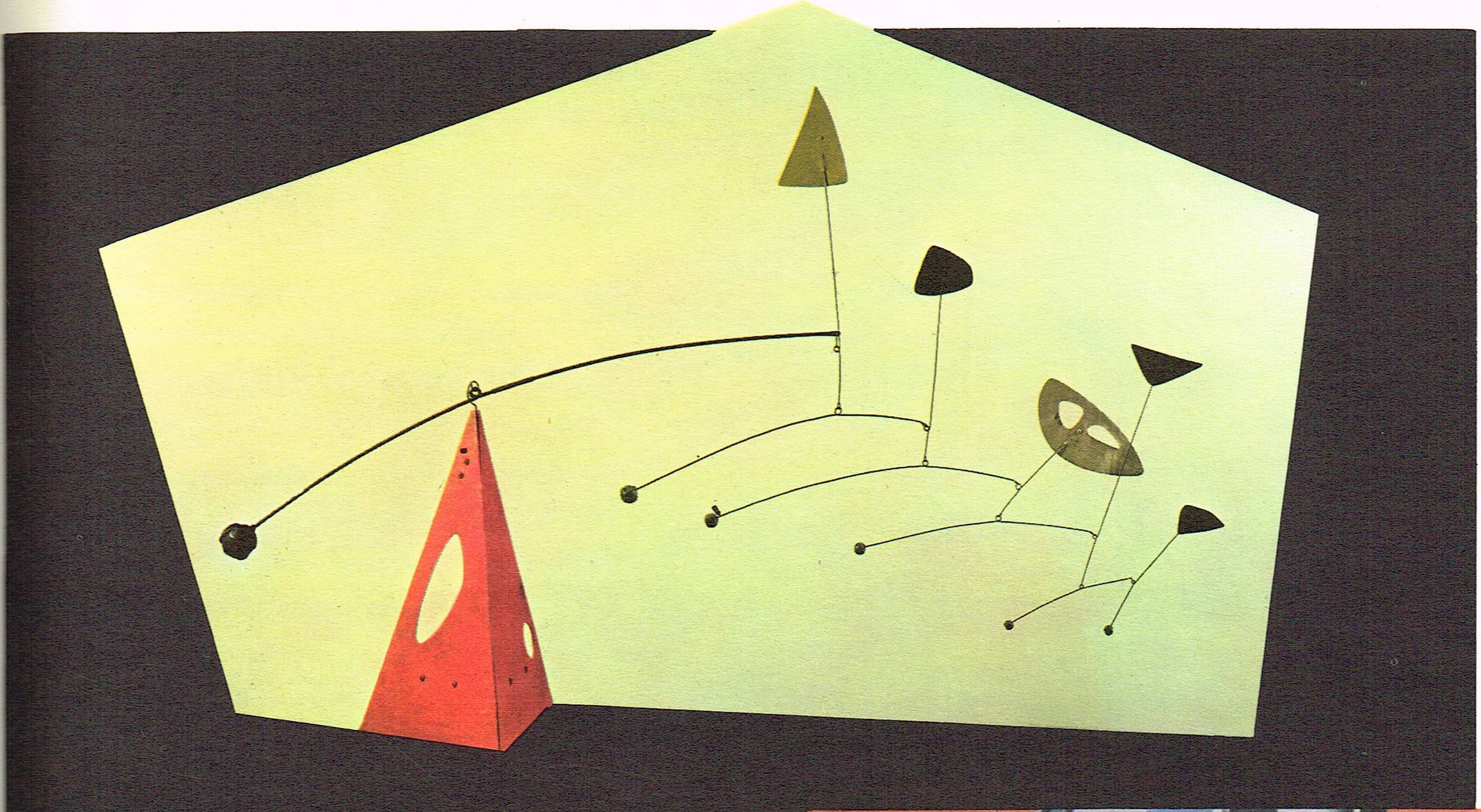
A une telle question, on peut donner plusieurs réponses plus ou moins élaborées.

Une première, simple et excellente, est la suivante : les accélérations sont verticales, l'obus devra donc décrire une trajectoire verticale (dans le référentiel Terre, qui est bien évidemment celui qu'il faut considérer) et finira par retomber dans la bouche à feu.

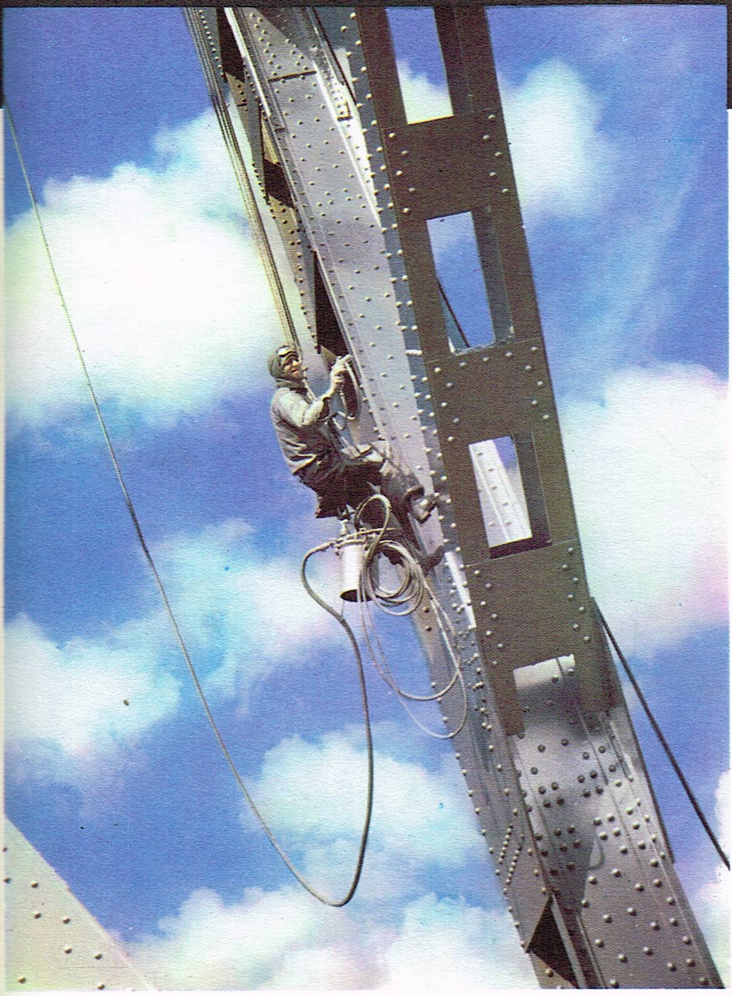
On a éludé ainsi le problème posé par la rotation de la Terre, ce qui nous amène à une seconde réponse, plus élaborée, mais aussi plus instructive : la durée du trajet est si petite que le mouvement de la pièce d'artillerie, rapporté maintenant au référentiel Terre-Étoiles, peut être tenu pour un mouvement rectiligne uniforme, et l'obus n'est soumis qu'à l'accélération de la gravitation terrestre, que l'on peut tenir pour constante. Alors, l'observateur du référentiel Terre-Étoiles se trouve vis-à-vis de cette expérience exactement dans la situation où se trouvait l'observateur terrestre qui regardait l'obus qu'un avion, en vol rectiligne horizontal et uniforme, tirait suivant sa propre verticale. L'obus retombera donc dans la bouche à feu.

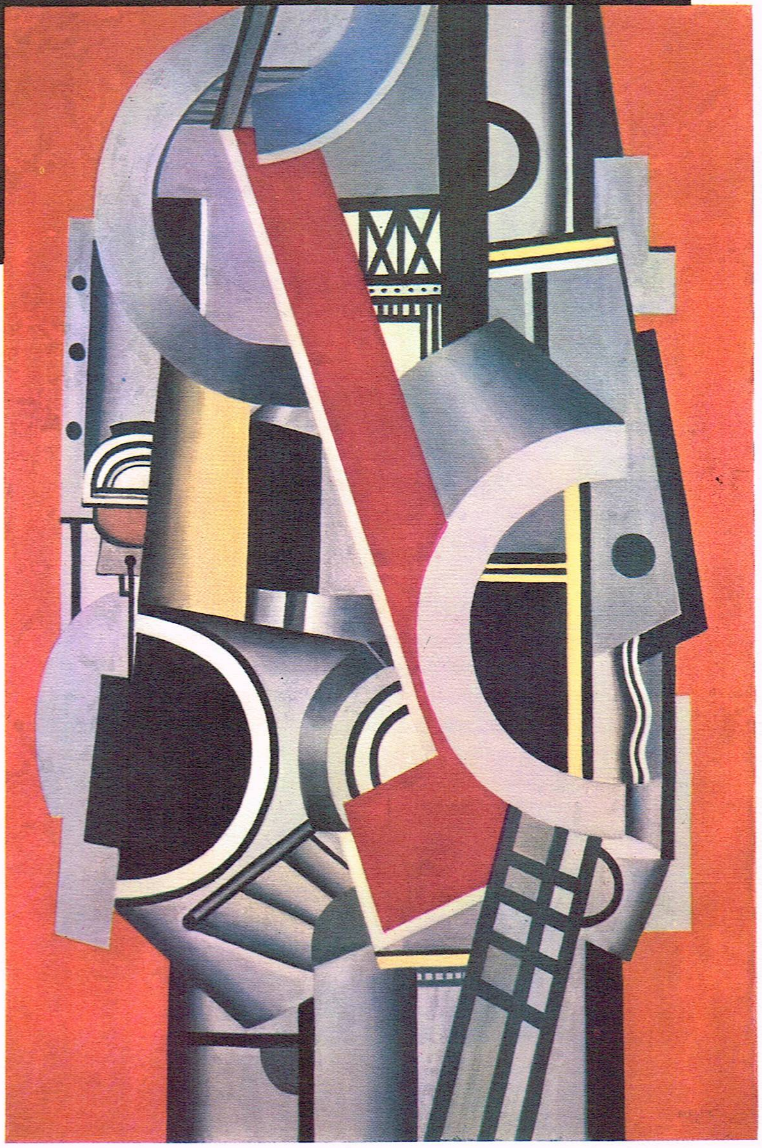
(1) Nous avons voulu montrer comment s'introduisait physiquement l'accélération de Coriolis. Bien entendu, les mathématiciens la retrouvent (et même l'ont trouvée bien avant que les expériences précises de Foucault l'aient mise en évidence) en dérivant la formule relative aux vitesses.

Phot. « l'Actualité française ».



« Red pyramid », mobile d'A. Calder. *Coll. Louis Carré, Paris.*

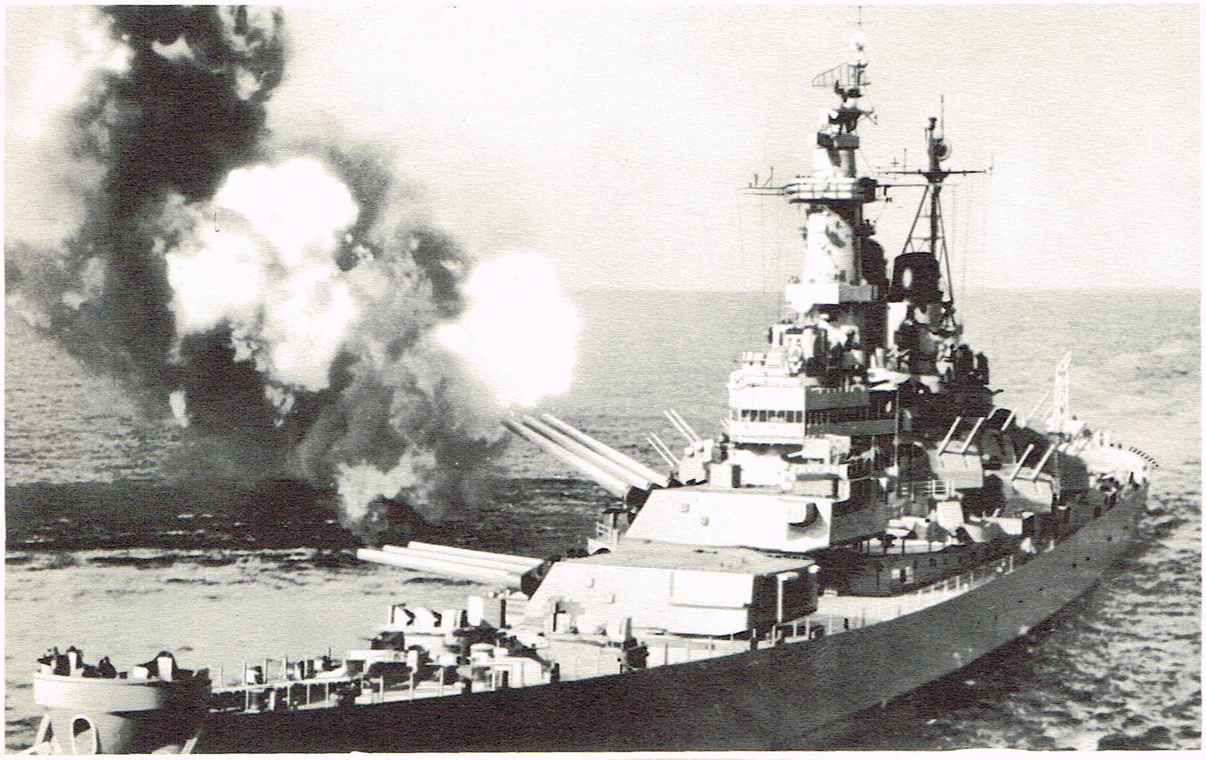




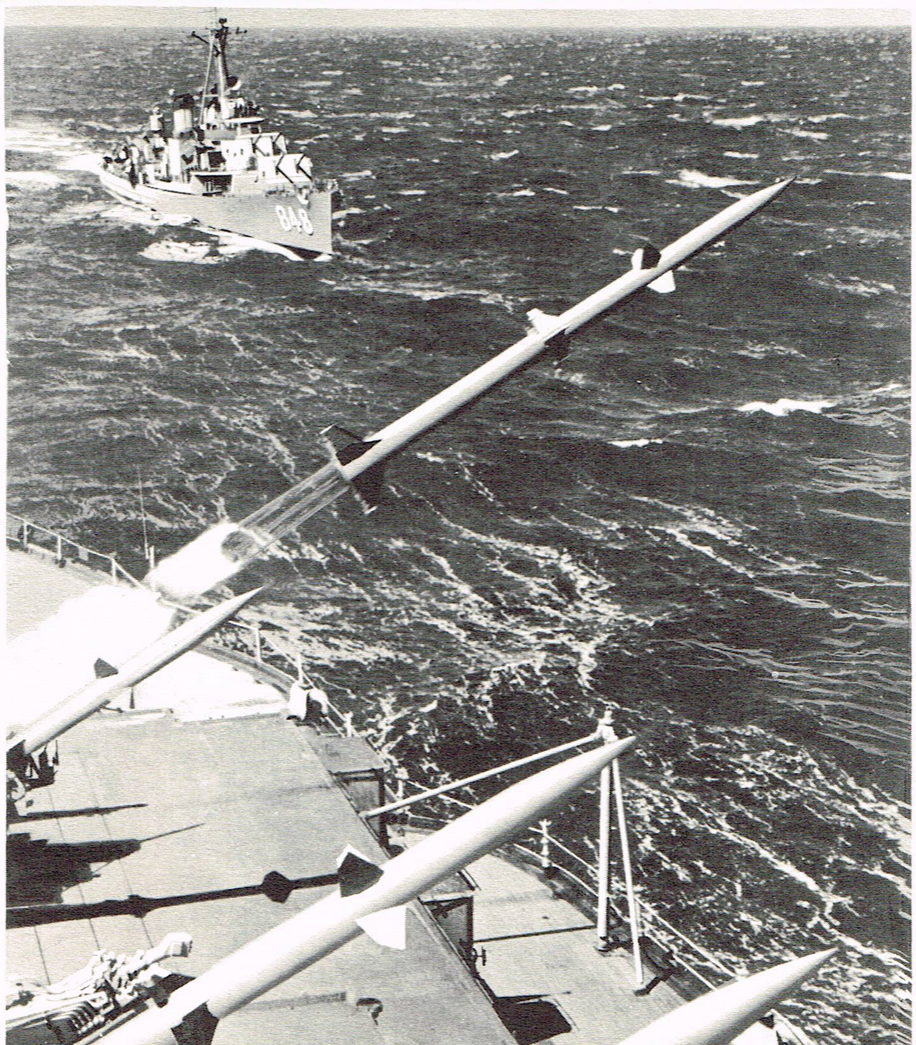
Réalité et fiction mécaniques.

Une imposante construction métallique aux U. S. A. *Doc. Esso.*

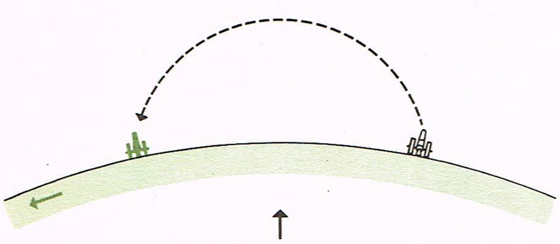
« Éléments mécaniques », par F. Léger. *Coll. Gourgaud, Paris. Phot. Giraudon.*

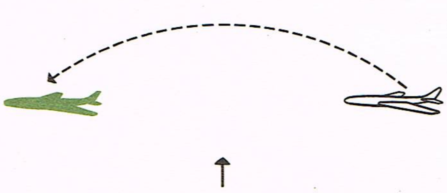


Un projectile à longue portée, qu'il soit tiré par un canon ou autopropulsé dans le plan méridien, devrait y rester constamment. Il est en fait, lui aussi, dévié vers l'est, et il faut en tenir compte dans les tirs précis. Jusqu'à ces dernières années on ne faisait intervenir cette déviation vers l'est que pour les seuls tirs de marine; à présent, elle intervient évidemment de façon importante dans les tirs des engins balistiques intercontinentaux.

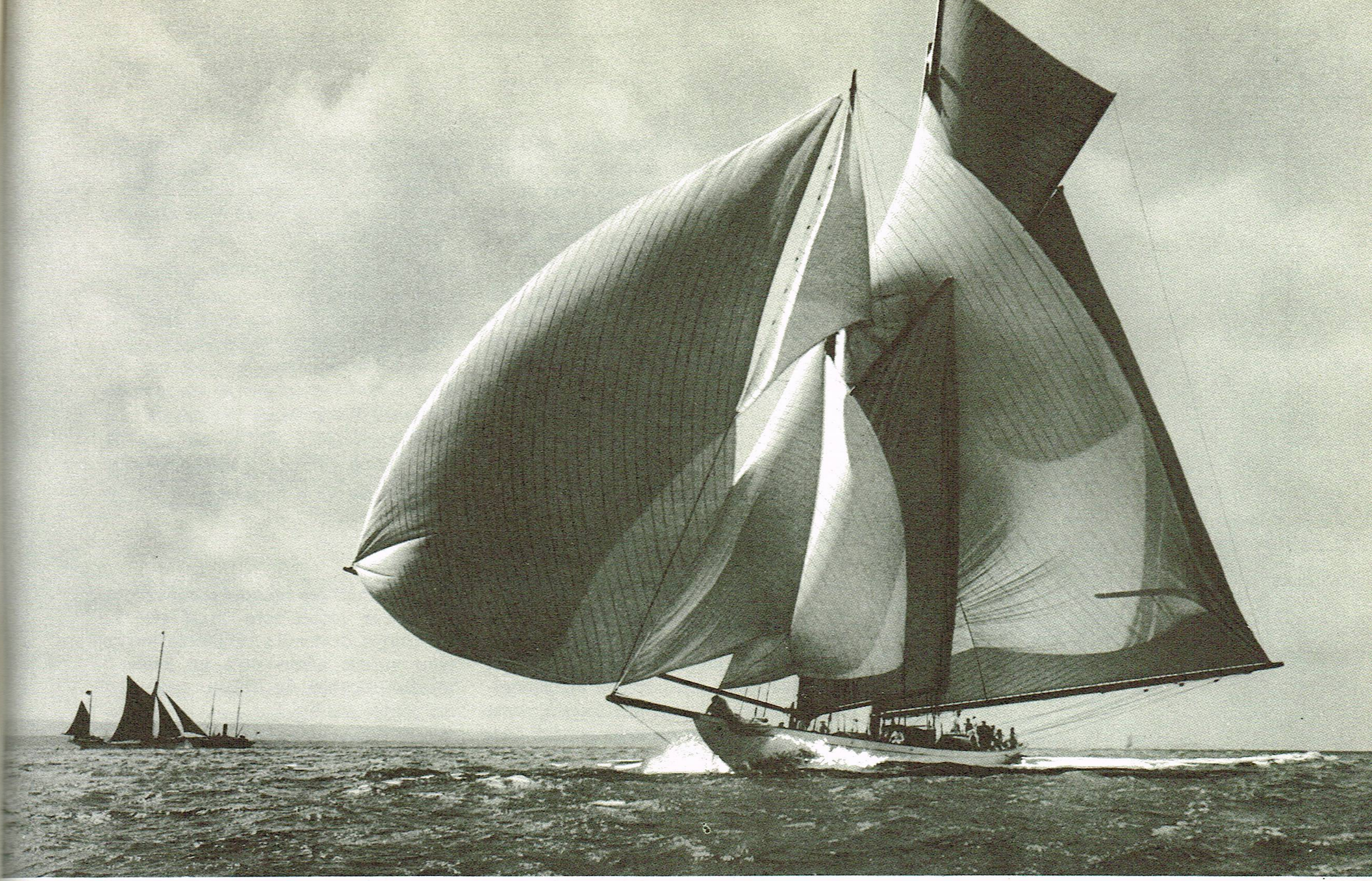


On pourrait encore donner d'autres réponses, plus élaborées, si l'on se refusait à assimiler le mouvement de la pièce à un mouvement rectiligne uniforme ou si l'on voulait tenir compte des conditions les plus fines de l'expérience (résistance de l'air, vent, autres actions en cause, irrégularité de l'âme du canon, etc.), ce qui serait possible au prix des calculs les plus compliqués et les plus ardus. Disons simplement ce que tout le monde sait d'ailleurs : qu'il vaut mieux s'éloigner après le départ du projectile, qui, de toute façon, ne retombera pas loin !





Phot. U. S. I. S.



Utiliser les actions naturelles pour diminuer l'effort musculaire.

les forces, causes et explication du mouvement

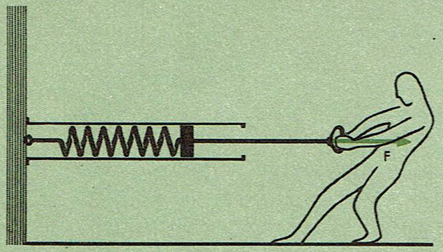
Les mouvements que nous venons de décrire et d'expliquer par l'accélération de gravitation terrestre sont des mouvements très particuliers. Le mobile y est inerte et l'espace vide. Comment expliquer le mouvement de l'engin spatial muni de réacteurs qu'un cosmonaute ou un système de télécommande peut mettre en marche à un instant déterminé ? Comment tenir compte de l'effet de la résistance de l'air sur son mouvement, ou sur le mouvement des graves ? Que va-t-il rester de notre explication lorsqu'il s'agira d'étudier le mouvement de la fusée qui a servi à lancer le satellite, et tout autre mouvement dont doit finalement s'occuper, par définition même, la science mécanique ?

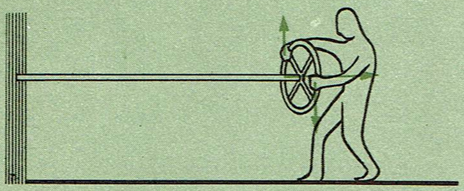
Seul subsistera le fait fondamental que le calcul des mouvements doit être effectué à partir de l'accélération. Avec le principe des conditions initiales, cela nous garantit une mise en équation correcte du problème. Mais, puisque l'accélération de gravitation terrestre n'explique pas tout, il nous faut chercher à connaître *a priori* l'accélération totale du mobile. Cette accélération doit être connue lorsque l'on connaît toutes les actions exercées sur le mobile.

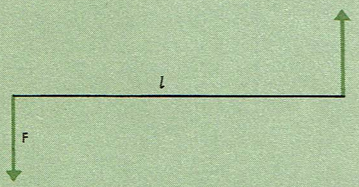
Pour analyser les actions qui s'exercent sur les corps matériels, faisons appel à des observations et à des expériences très communes.

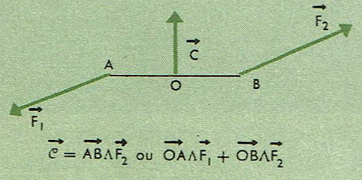
Les êtres animés ont la possibilité d'exercer des actions sur les objets qui les entourent. L'homme a dû savoir très tôt analyser ces actions, les utiliser et les perfectionner pour atteindre le but qu'il se fixait. Notre propre déplacement, celui d'un objet par l'homme, ou encore la déformation qu'il faut lui imposer pour lui donner une forme plus utile, sont obtenus au prix d'un effort musculaire plus ou moins considérable. Rapidement, l'homme a su utiliser des actions naturelles pour diminuer l'effort qu'ont à fournir ses muscles, ou ceux de ses animaux domestiques : la pesanteur pour déplacer un bloc de pierre le long d'une colline, l'action du vent pour faire tourner des éoliennes de moulin, etc. Puis il a su construire des outils et des machines, qui d'abord ne lui ont servi qu'à utiliser l'effort musculaire avec un meilleur rendement, et ensuite ont travaillé pour lui en empruntant leur énergie

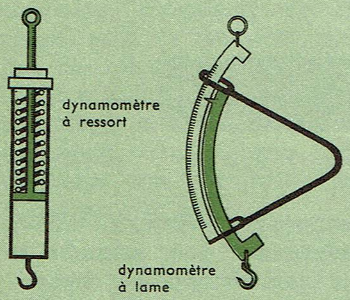
Phot. Beken and Son.

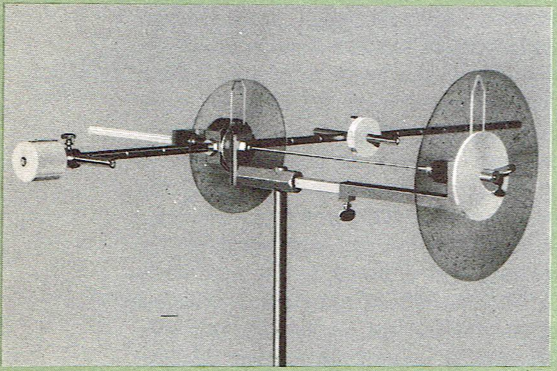












à d'autres sources que les muscles. Toutes ces actions mettent en jeu des forces. Alors, généralement, nous appellerons « force » toute grandeur physique susceptible de déformer un corps ou de le mettre en mouvement.

Examinons d'abord, à l'aide d'expériences très simples, le premier aspect de cette notion de force, grandeur susceptible de provoquer la déformation d'un corps au repos.

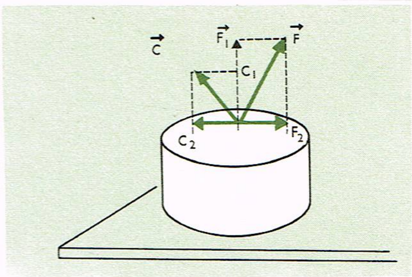
L'homme de la figure exerce sur le ressort un effort musculaire dont l'effet est d'allonger le ressort. Cet effort s'exerce dans une direction et dans un sens bien déterminés. Il a, de plus, une intensité plus ou moins grande suivant l'allongement désiré. La force est une grandeur vectorielle.

Et l'on établira entre la force et l'allongement une relation de cause à effet, c'est-à-dire que, inversement, chaque fois que nous constaterons un allongement, nous dirons qu'une force s'exerce sur notre ressort, qui nous permet donc de déceler des forces.

Pour se rendre compte de l'intensité de la force, le plus simple est alors de mesurer la cause par son effet, la force par l'allongement : à un allongement double, triple, nous poserons en principe qu'il lui correspond une force double, triple. La force devient ainsi une grandeur vectorielle mesurable dont l'intensité pourra être caractérisée par un nombre, une fois que l'on aura fait choix de l'unité de force qui est la dyne dans le système C. G. S., le newton dans le système M. K. S. A. (Mètre, Kilogramme, Seconde, Ampère). Le ressort peut alors être étalonné, si on lui adjoint une graduation. Il deviendra un appareil à mesurer les forces, un dynamomètre.

Une longue barre métallique, de faible section, est encastrée par une de ses extrémités. Son autre extrémité est munie d'un volant. Un poids (v. phot. ci-contre) ou un effort musculaire exercé sur le volant provoque une torsion de la barre, déformation que l'on peut mesurer par l'angle dont a tourné le volant.

Cet effort, que l'on appelle un **couple,** est caractérisé par la direction de l'axe autour duquel a tourné le volant, le sens de rotation du volant et l'intensité de cet effort. Ce couple peut être représenté par un vecteur porté par l'axe de rotation, à condition de faire une convention de sens : si l'opérateur de la figure (ci-contre) voit tourner le volant dans le



Ici, la force n'est pas seulement une traction ($\vec{F_1}$), mais comporte un « cisaillement » ($\vec{F_2}$); le couple n'est pas seulement une « torsion » ($\vec{\mathcal{C}_1}$), mais aussi un « fléchissement » ($\vec{\mathcal{C}_2}$).

sens contraire des aiguilles d'une montre, le vecteur couple sera dirigé vers lui. Ce couple deviendra mesurable lorsque nous aurons admis qu'à un angle de rotation double du volant correspond un couple double.

Le couple n'est pas une notion nouvelle par rapport à la force. Chaque main de l'opérateur exerce sur le volant une force; ces deux forces sont parallèles, ont des sens opposés et la même intensité. L'intensité du couple est le produit de l'une de ces forces F par le diamètre du volant *l*, mais on ne se préoccupe pas de savoir si le couple provient de deux forces intenses et rapprochées, ou de deux forces faibles et éloignées. Le vecteur représentant le couple est, en fait, un produit vectoriel.

De la même manière que pour le ressort, chaque fois que l'on aura pu déceler une rotation de l'extrémité libre de la barre, on dira qu'un couple s'exerce sur elle. La barre, étalonnée par l'adjonction d'une graduation, deviendra un appareil à déceler et à mesurer les couples. On utilise plutôt en physique un fil de torsion qui, muni d'un miroir renvoyant l'image d'une fente lumineuse sur une règle graduée, constitue un instrument très précis et très sensible : le pendule de torsion.

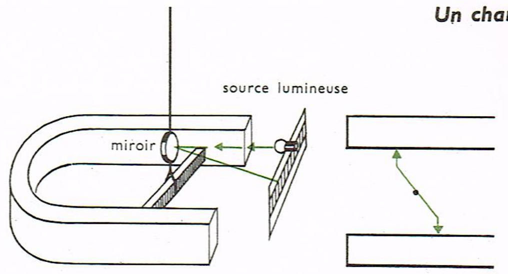
mise en évidence et mesures générales des forces

Ces deux expériences fondamentales permettront de lier d'une façon générale les notions de force et de déformation. Ainsi, une barre soumise en même temps à une force de traction et à un couple de torsion, se déforme d'une façon

complexe, qui apparaît comme la superposition des deux déformations précédentes. D'une façon générale, on rendra compte de la déformation d'un objet à l'aide d'une force et d'un couple.

Toute déformation d'un corps au repos met en évidence la présence de forces appliquées à ce corps. Inversement, une ou plusieurs forces appliquées à ce corps lui font toujours subir une déformation. Si nous ne la voyons pas dans certains cas où l'objet est trop solide ou les forces trop faibles, c'est que nous ne regardons pas les choses d'assez près; il conviendra de faire des observations plus précises.

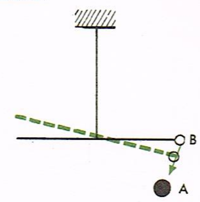
Un champ magnétique exerce un couple sur un barreau aimanté.

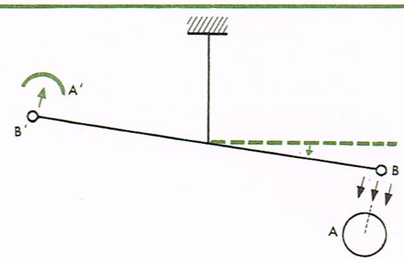


Un barreau aimanté placé dans l'entrefer d'un aimant en fer à cheval décèle l'existence de forces magnétiques se traduisant par un couple.

Les corps chargés d'électricité s'attirent ou se repoussent.

Le corps chargé A exerce sur le corps chargé B une force de nature électrique qui peut être une attraction ou une répulsion.





Le corps A maintenu fixe exerce sur le corps B une force $\vec{F}$ que le pendule de torsion permet de mesurer.

quelques catégories de forces

Les deux premières forces (ci-contre), de nature magnétique ou électrique, seront étudiées dans un autre chapitre de cet ouvrage. Mais le pendule de torsion permet de déceler et de mesurer des forces sur lesquelles nous allons réfléchir davantage.

même deux corps neutres s'attirent

La célèbre expérience de Cavendish (1731-1810) met en évidence la force d'attraction que deux corps quelconques, même neutres, exercent l'un sur l'autre.

Des expériences variées, qu'il est facile d'imaginer et dans le détail desquelles nous n'entrerons pas, fournissent les résultats suivants :

1° Pour deux corps A et B déterminés, la force F est inversement proportionnelle au carré de la distance r qui les sépare. Cela reste vrai quelle que soit leur position respective; le coefficient de proportionnalité ne dépend que des deux corps en présence. On peut écrire

$$F = \frac{K(A, B)}{r^2};$$

2° Pour un corps B déterminé, la force d'attraction caractérise le pouvoir d'attraction de A. Le rapport des forces F_{A_1}, F_{A_2} exercées par deux corps A_1 et A_2 sur B est indépendant du corps B, du lieu et de l'époque de l'expérience. Il peut donc servir à définir le rapport des pouvoirs attractifs de ces corps, pouvoir que l'on caractérisera par des nombres, les **masses gravitationnelles.** On définira le rapport entre les masses m_{A_1} et m_{A_2} des corps A_1 et A_2 comme étant le rapport entre F_{A_1} et F_{A_2}. La masse devient ainsi une grandeur mesurable, une fois que l'on aura fait choix de l'unité de masse, qui est la masse du kilogramme-étalon déposé au pavillon de Breteuil, à Sèvres. La force F exercée par A sur B est ainsi proportionnelle à la masse de A, le coefficient de proportionnalité ne dépendant plus que du corps B. On peut écrire

$$F = K_1(B)\,\frac{m_A}{r^2};$$

3° Pour un corps A déterminé, la force d'attraction change avec le corps B, et un raisonnement analogue au précédent montre que la force d'attraction est proportionnelle aussi à la masse de B.

On peut écrire $F = k\,\dfrac{m_A m_B}{r^2}$.

la constante de la gravitation

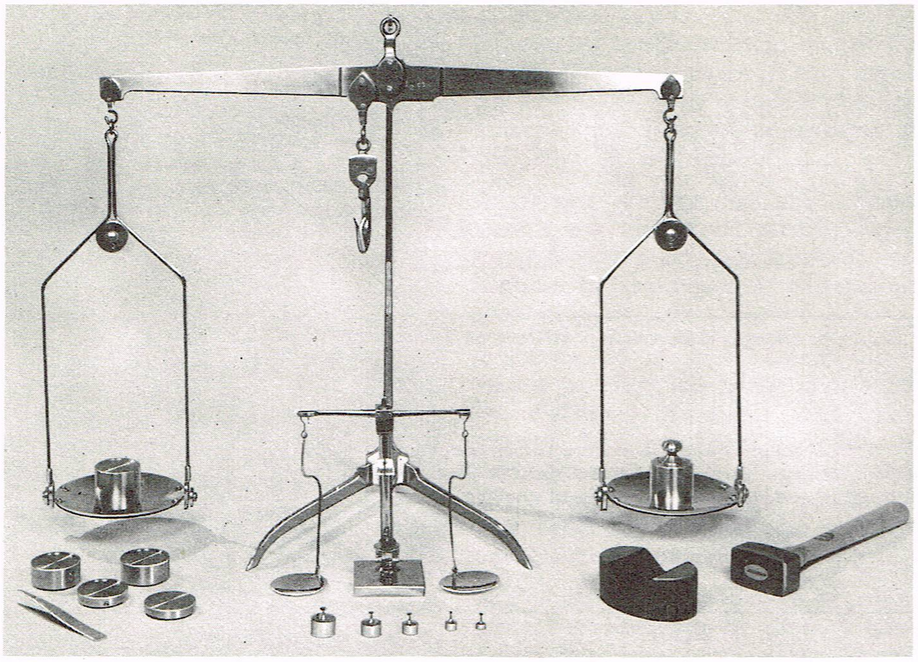
Le coefficient de proportionnalité k ne dépend plus alors ni du corps A, ni du corps B, ni du lieu, ni de l'époque de l'expérience. C'est une constante universelle, que l'on appelle la « constante de la gravitation ». Sa valeur numérique dans un certain système d'unités est égale à l'intensité de la force qu'exerce sur une masse-unité une autre masse-unité située à l'unité de distance. Les expériences qui la mesurent font intervenir les oscillations du pendule de torsion, dont on mesure la période, et fournissent la valeur

$$k = 6{,}67 . 10^{-8} \text{ C. G. S.}$$
$$(6{,}67 . 10^{-11} \text{ M. K. S. A.}),$$

avec une précision, pas très bonne, de l'ordre du millième. Les forces d'attraction sont faibles : une masse de 1 g, placée à 1 cm d'une autre masse égale, exerce sur cette dernière une force de 0,000 000 066 7 dyne, et la dyne ne vaut qu'environ un millionième de la force que nous appelons habituellement un « kilo »! En revanche, elles deviennent considérables lorsque les masses en présence sont celles du Soleil, des planètes ou des étoiles. Ainsi, l'attraction exercée par le Soleil sur la Terre est de 3,5 milliards de milliards de tonnes.

le poids

Des mesures variées montrent que poids et masse gravitationnelle sont proportionnels : le rapport des poids de deux corps est égal au rapport de leurs masses, indépendamment du lieu et de l'époque



Nécessaire portatif pour le contrôle des poids du commerce.

Deux masses égales ont le même poids en tout lieu.

De l'équilibre de la balance on déduit l'égalité des masses.

des mesures. Ce qu'il convient alors de connaître, c'est, en chaque lieu et à chaque époque, le poids de la masse-unité. En promenant le kilogramme-étalon attaché à un dynamomètre, on pourra construire le champ de la force de pesanteur $\vec{G}$ s'exerçant sur la masse-unité. Des instruments (pendule de Lejay,...) permettent de mesurer et de connaître le champ $\vec{G}$, du moins à la surface de la Terre; il est constant dans une portion d'espace pas trop étendue, mais varie avec le lieu et aussi, quoique de manière imperceptible, avec l'époque de la mesure (l'étude de ce champ constitue à elle seule toute une science, la gravimétrie, qui nous permet d'étudier la répartition des masses à l'intérieur de la Terre et de déceler les nappes de pétrole ou d'eau).

C'est la proportionnalité entre le poids en n'importe quel lieu et la masse, qui, elle, est indépendante du lieu, qui permet de se servir d'une balance pour mesurer des masses.

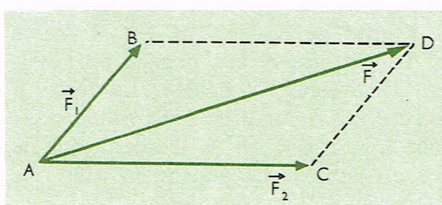
Nous aurons à revenir sur la notion de poids lorsque nous aborderons une partie « explicative ».

l'équilibre des forces

Maintenant que nous avons appris à déceler et à mesurer les forces agissant sur les corps au repos, il s'agit d'expérimenter sur ces forces. L'étude des forces assurant l'équilibre des corps constitue la **statique.** Nous nous contenterons d'en donner les résultats.

1° La réaction est toujours égale à l'action; c'est le principe de l'action et de la réaction.

2° Il revient au même d'appliquer au corps A les deux forces $\vec{F}_1$ et $\vec{F}_2$, ou leur résultante $\vec{F}$ obtenue par addition géométrique.



3° Enfin, les expériences de statique montrent que si un corps ponctuel est au repos, la résultante de toutes les forces qui lui sont appliquées est nulle. Inversement, si un corps ponctuel est abandonné sans vitesse en un point où la résultante de toutes les forces appliquées est nulle, ce corps demeurera au repos.

la force est ce qui déplace

Toute poussée exercée sur un corps immobile le met en mouvement; toute poussée exercée sur un corps déjà mobile en modifie le mouvement; alors la force apparaît comme la grandeur susceptible de créer ou de modifier un mouvement. Pour créer un mouvement, on utilise des forces de toute nature.

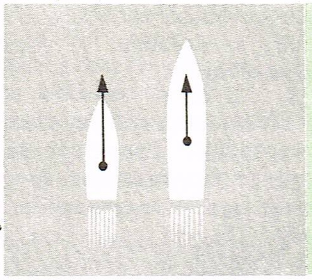


L'effort musculaire du frondeur provoque un mouvement circulaire de plus en plus rapide de la pierre.

L'arc courbé par la force musculaire du tireur tend à reprendre sa forme primitive et exerce une force sur la flèche.

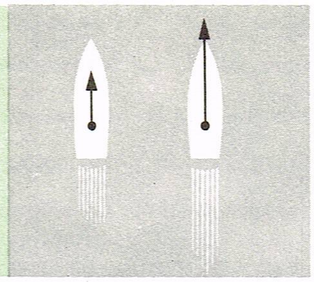


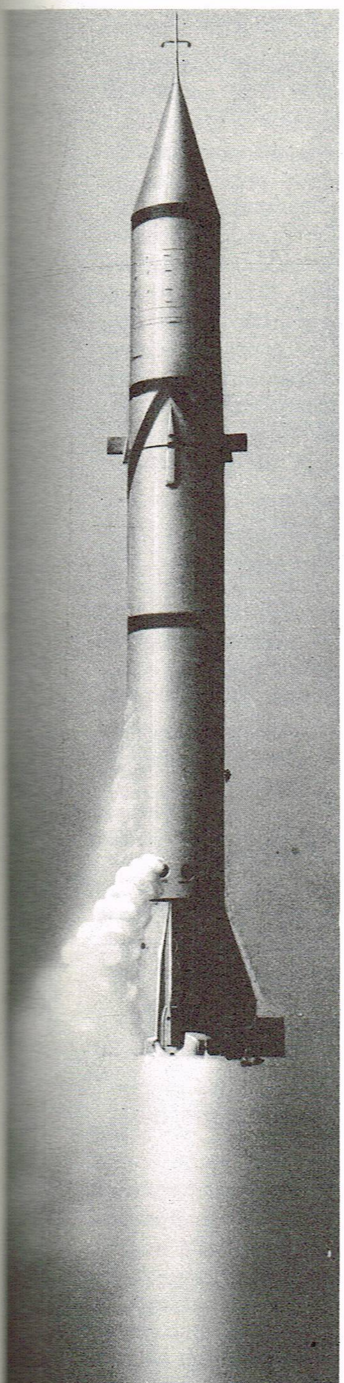
Phot. Larousse, Giraudon.



Pour une même poussée, l'accélération est inversement proportionnelle à la masse.

Pour une même masse, l'accélération est proportionnelle à la poussée.





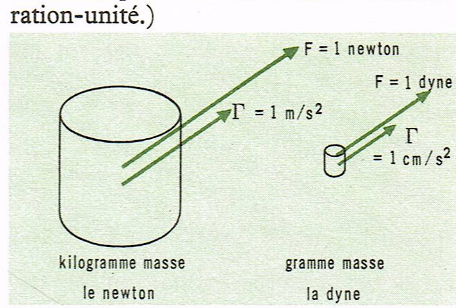
Les gaz formés par la combustion du propergol exercent une forte poussée sur la fusée.

forces et mouvements

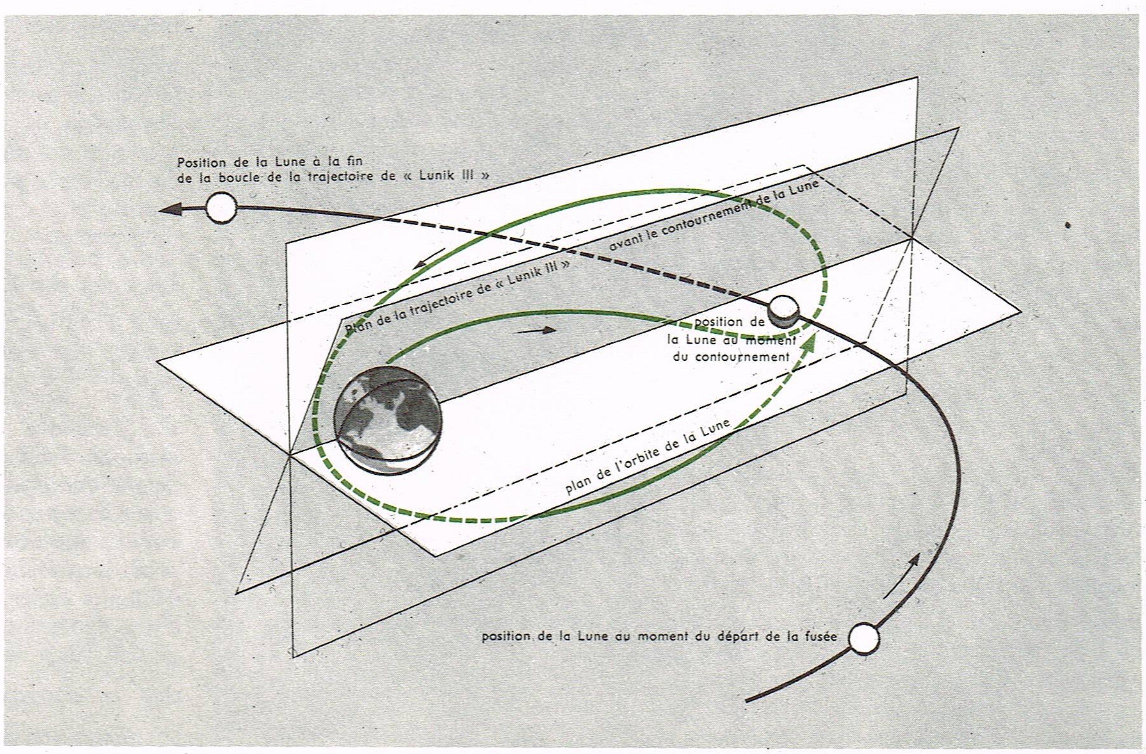
L'étude des forces en relation avec le mouvement créé ou modifié du corps sur lequel elles s'exercent constitue la **dynamique.** Tout d'abord, le principe des conditions initiales, que nous n'avons énoncé antérieurement que pour des mouvements particuliers, sera étendu à tous les mouvements. Pour préciser la relation entre forces et mouvement, nous nous adresserons à des mouvements produits par l'action de forces connues, et c'est l'accélération qui va être l'objet de nos mesures puisqu'elle a déjà servi pour des mouvements particuliers et qu'elle convient au cadre du principe des conditions initiales.

A Cap Canaveral, par exemple, chaque départ de fusée donne lieu à des mouvements particulièrement spectaculaires. Les ingénieurs de la base, s'ils voulaient bien se mettre à notre disposition, pourraient nous dire quelle est la poussée exercée au départ d'une fusée par la combustion du propergol utilisé. Ils pourraient nous dire quelle est la masse de la fusée. Nous leur demanderions aussi de mesurer l'accélération de la fusée. Puis nous assisterions à des départs de fusées de différents types et rassemblerions (ci-contre) les résultats obtenus.

Donc, l'accélération est bien liée à la force, mais par l'intermédiaire de la masse; l'accélération est proportionnelle à la force et inversement proportionnelle à la masse : $\vec{\gamma} = \frac{\vec{F}}{m}$. On écrira $\vec{F} = m\vec{\gamma}$ et on précisera la relation entre force et mouvement en disant que la force $\vec{F}$ communique à la masse m l'accélération $\vec{\gamma}$, ces trois quantités étant reliées par la relation $\vec{F} = m\vec{\gamma}$. (Le coefficient de proportionnalité a été pris égal à l'unité, car, dans tout système d'unités, la force-unité est la force qui communique à la masse-unité l'accélération-unité.)



La formule $\vec{F} = m\vec{\gamma}$ que nous avons



« Lunik III » est attiré à la fois par la Terre et par la Lune.

Phot. U. S. I. S., Bureau soviétique d'information.

dégagée d'un cas très particulier est générale. Dans tous les mouvements qui ont lieu sous l'action d'une force connue, on observe que l'accélération s'obtient en divisant cette force par la masse. Inversement, tous les mouvements calculés à l'aide de cette accélération et des conditions initiales peuvent être effectivement réalisés. Cette vérification, qui fait appel à des observations et à des mesures, ne donne évidemment que des résultats approchés. Tout ce que nous pouvons dire, c'est qu'aucune expérience faite sur des mouvements dus à des forces connues ne décèle de différence entre $\vec{F}$ et $m\vec{\gamma}$. Mais nous n'en avons pas pour autant terminé avec cette formule, que nous pensons devoir être générale, que nous désirons rendre générale et que nous appelons dès à présent **formule fondamentale de la dynamique.** En effet, bien des points restent à examiner.

D'abord les mouvements décrits précédemment, mouvements des graves et mouvements des satellites, qui ont été expliqués par l'accélération, demandent à présent à être expliqués par la force, ce qui est aisé.

$\vec{F} = m\vec{\gamma}$ pour les mouvements des graves

En ce qui concerne les mouvements des graves, nous avons vu d'une part que tout corps lâché à la surface de la Terre possède une accélération $\vec{g}$ constante (dans un domaine pas trop étendu); d'autre part, qu'il est soumis à une force, son poids $\vec{P} = m\vec{G}$. Nous pouvons expliquer aussi bien le mouvement des graves par l'accélération $\vec{g}$ ou par la force $\vec{P}$. La formule fondamentale permet, en effet, d'identifier le champ d'accélération dû à la pesanteur au champ de force que cette même pesanteur exerce sur la masse-unité.

$\vec{F} = m\vec{\gamma}$ pour les mouvements des satellites

En ce qui concerne les mouvements des satellites artificiels de la Terre, nous n'avons encore à notre disposition que leur accélération et leur masse. Nous avons donc deux termes de la relation fondamentale $\vec{F} = m\vec{\gamma}$ et nous pouvons en tirer le troisième. Nous dirons donc que le satellite est soumis à une force, dirigée vers le centre de la Terre, égale au produit de sa masse m par son accélération, c'est-à-dire, d'après nos notations, à $\frac{m\mu}{r^2}$. Ce mouvement peut donc s'expliquer, là encore, en termes de force, $\frac{m\mu}{r^2}$, ou d'accélération, $\frac{\mu}{r^2}$. Nous substituons donc au champ des accélérations de la gravitation terrestre un champ de forces; un champ de forces entoure la Terre, il est central, proportionnel à la masse d'épreuve et inversement proportionnel au carré de la distance de cette masse au centre de la Terre.

$\vec{F} = m\vec{\gamma}$ pour tout mouvement

Nous avons acquis l'idée, fondée sur des expériences, que la loi $\vec{F} = m\vec{\gamma}$ va régir tous les mouvements. Ce qu'il nous faut, c'est avoir une idée précise sur le contenu des différents termes intervenant dans cette relation.

Nous ne reviendrons pas sur la notion de temps, qui intervient dans la mesure d'accélération.

Nous avons rencontré la masse à différentes reprises, sous des aspects différents. Tout d'abord, c'est la **masse gravitationnelle,** que nous avons rencontrée dans l'expérience de Cavendish (1).

Puis, lorsque nous avons attaché un corps à un dynamomètre et que nous avons mis en évidence le poids de ce corps, nous avons affirmé que ce poids était proportionnel à la masse gravitationnelle. En réalité, le fait que le rapport des poids de deux corps est constant permet de définir la **masse pesante** d'un corps. L'identité de la masse pesante et de la masse gravitationnelle est un résultat expérimental.

Ensuite, en dynamique, la première fois que nous avons rencontré $\vec{F} = m\vec{\gamma}$, au départ des fusées de Cap Canaveral, la masse est intervenue sous un aspect nouveau. Pour une même force appliquée, plus la masse est élevée, plus faible est l'accélération, c'est-à-dire plus difficile est la mise en mouvement du corps. La masse caractérise ici l'**inertie** du corps, c'est-à-dire son pouvoir de résistance à sa mise en mouvement ou à la modification de son mouvement. Il s'agit ici d'une **masse inerte.**

(1) En toute rigueur d'ailleurs, il y aurait eu lieu de définir une masse gravitationnelle attirante, celle du corps A de notre expérience, et une masse gravitationnelle attirée (celle du corps B). Mais l'échange des corps d'épreuve A et B, ou le principe de l'égalité de l'action et de la réaction, prouve (toujours expérimentalement) qu'il n'y a pour un corps qu'une seule masse gravitationnelle.

L'explication du mouvement des satellites en termes de force d'attraction par la Terre implique l'identification de la masse inerte du satellite avec sa masse gravitationnelle. D'autre part, l'explication par la force du mouvement des graves implique l'identification de la masse inerte avec la masse pesante.

En résumé, en mécanique classique, toute difficulté relative à la notion de masse est éliminée, mais il faut bien comprendre que cela est un résultat expérimental : aucune expérience du domaine de la mécanique classique, effectuée avec la plus grande précision que nous pouvons atteindre à ce jour, ne permet de déceler des différences entre diverses sortes de masses, ni des variations de masse au cours du temps ou lors d'un mouvement. Pour nous donc, la masse est indestructible, invariable dans le temps et dans l'espace, indépendante de l'état de mouvement ou de repos du corps.

En ce qui concerne la force, nous avons vu quelle confiance on peut attribuer à la formule $\vec{F} = m\vec{\gamma}$ lorsque la force est supposée connue. Mais comment être sûr que l'on connaît bien toutes les forces? Simplement en affirmant que cette formule $\vec{F} = m\vec{\gamma}$ est toujours vraie, en mesurant des accélérations dans des catégories très diverses de mouvements et en déduisant à l'aide de $\vec{F} = m\vec{\gamma}$ la force à partir de l'accélération. C'est l'extension de ce que nous avons fait à propos des mouvements des satellites de la Terre, pour lesquels nous avons inventé le champ des forces de gravitation terrestre.

Insistons sur l'utilisation de cette formule $\vec{F} = m\vec{\gamma}$, qui présente un double aspect, l'un expérimental, l'autre théorique.

1° L'existence d'une force peut être mise en évidence par une expérience de statique ou de dynamique. S'il s'agit d'une expérience de statique, la mesure au dynamomètre indiquera en chaque point la présence d'une force, qui peut d'ailleurs varier avec le temps. S'il s'agit d'une expérience de dynamique, le fait qu'une masse m a une certaine accélération $\vec{\gamma}$ traduira le fait qu'une force ($\vec{F} = m\vec{\gamma}$) s'exerce sur elle. Des mesures

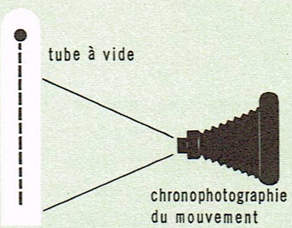
d'accélération en différents points, à différents instants et pour des passages du mobile avec des vitesses variées mettront en évidence l'existence de forces qui pourront dépendre de la position du mobile, du temps et de sa vitesse. Pour étudier ces forces, l'expérimentateur choisira des conditions favorables. Souvent il pourra décomposer $\vec{F}$ en forces élémentaires, plus faciles à étudier. L'expérience montre alors, en dynamique comme en statique, que les forces s'ajoutent suivant la règle du parallélogramme. C'est ce qu'on appelle le « principe » de **superposition** des actions, ou d'indépendance des actions : si $\vec{F}_1$ appliquée seule procure au mobile une accélération $\vec{\gamma}_1$, si $\vec{F}_2$ appliquée seule procure au mobile une accélération $\vec{\gamma}_2$, alors la résultante $\vec{F} = \vec{F}_1 + \vec{F}_2$ de ces forces lui procure l'accélération $\vec{\gamma} = \vec{\gamma}_1 + \vec{\gamma}_2$. Les « actions » se superposent purement et simplement, chacune d'elles continuant à agir comme si elle était seule.

2° L'analyse expérimentale des forces ayant été ainsi faite, la loi fondamentale sera utilisée dans l'autre sens, le premier membre étant à présent connu. La connaissance de l'accélération et des conditions initiales permettra de calculer théoriquement tous les mouvements possibles en fonction des conditions initiales. C'est dans la confrontation des résultats théoriques avec des expériences faites après coup et faisant intervenir toutes les forces introduites que se trouve la justification du processus scientifique utilisé.

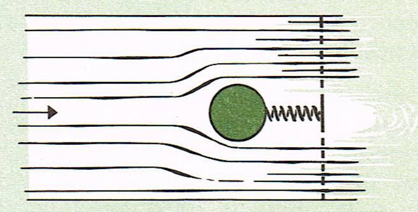
Illustrons ces considérations par un exemple concret : toujours celui de la chute des graves.

La formule du mathématicien, avec laquelle l'expérimentateur est entièrement d'accord, montre qu'il existe une vitesse limite, correspondant à une résistance de l'air exactement opposée au poids. Mais l'utilisation diffère : pour le mathématicien, cette limite n'est jamais atteinte; pour le physicien, elle l'est « pratiquement » au bout d'un temps qui peut être très court. Ainsi, il considérera qu'une gouttelette d'eau descend dans l'atmosphère avec une vitesse uniforme.

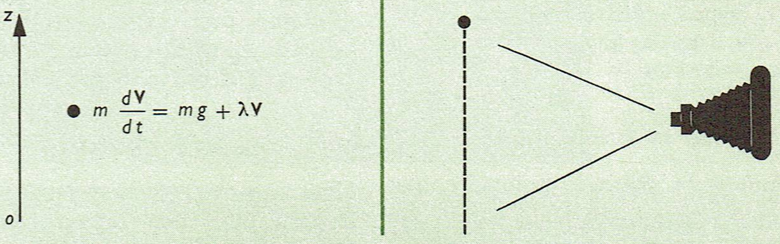
Dans la spectaculaire expérience des sauts avec parachute à ouverture retardée, on peut considérer pratiquement que la chute libre, qui a lieu pendant un temps



La chute verticale d'un corps dans le vide décèle le poids du corps $\vec{\Gamma} \to \frac{\vec{P}}{m}$.



Des essais en soufflerie permettent d'étudier la force due à la résistance de l'air $\vec{R} = \lambda\vec{V}$.



Le mathématicien calcule le mouvement de chute dans l'air. *A droite,* l'expérimentateur étudie la chute dans l'air. Les mouvements qu'il observe sont bien ceux qui ont été prévus par le calcul du mathématicien.



Saut à ouverture retardée.

Phot. Usis.

pouvant dépasser la minute, se fait à vitesse constante de l'ordre de 500 km/h, avant que soit déclenchée l'ouverture du parachute à quelque 300 m du sol.

L'étude faite en soufflerie du coefficient λ montre que, de deux corps de forme identique, l'effet de la résistance de l'air est beaucoup plus fort sur celui qui a la masse la plus faible. Ainsi, de deux sphères de même rayon, l'une en plomb, l'autre en bois, qui, dans le vide, tomberaient de la même façon, c'est la moins massive, celle qui est en bois, qui sera la plus retardée par la résistance de l'air.

mais tout cela est-il bien sérieux ?

Un humoriste pourrait prétendre que notre façon de traiter la relation $\vec{F} = m\vec{\gamma}$ n'est pas très sérieuse et qu'elle la vide de toute signification. « Evidemment, nous dira-t-on, la loi fondamentale est certainement vraie, avec votre façon d'opérer : si vous décelez quelque différence entre les mouvements calculés à l'aide de cette formule et les mouvements observés, vous n'avez qu'à prétendre qu'il existe des forces non décelées, les analyser par la différence entre le calcul et l'observation, les faire entrer alors dans le terme $\vec{F}$ de la formule, et vous êtes ainsi assurés que tout ira très bien. » En fait, c'est ainsi que les choses se présentent, mais ce qui n'aurait pas été sérieux, c'est de faire fi du bon sens, d'écrire $\vec{F} = m\vec{\gamma}$ dans des conditions arbitraires, ce qui aurait amené à introduire dans $\vec{F}$ des termes totalement dénués de toute signification physique, et, finalement, aurait rendu la loi fondamentale inutilisable. Ce qu'il ne faut pas perdre de vue, c'est le caractère expérimental de la mécanique et le fait que la construction de la loi fondamentale a été faite à partir de mouvements où la force était connue. Les différentes notions introduites, de temps, de masse, de force sous ses deux aspects, ainsi que les « principes » énoncés, viennent converger dans la loi fondamentale $\vec{F} = m\vec{\gamma}$. Les notions et les principes de la mécanique sont difficiles, mais la loi fondamentale est très simple; cela vaut mieux pour l'utilisateur.

Nous avons d'ailleurs l'idée très nette que dans l'état actuel de la science les forces ont été suffisamment analysées pour qu'il n'y ait pas de catégorie importante qui nous ait échappé. C'est donc à un type de force connu que l'on doit raisonnablement faire appel si nous décelons une différence entre les calculs et l'observation.

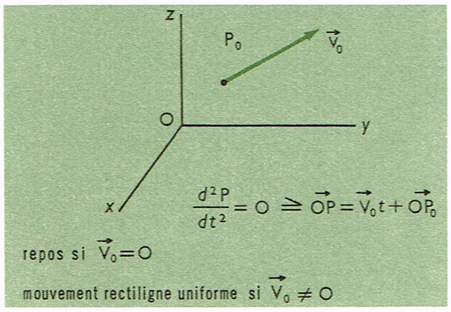
La célèbre aventure de Le Verrier (1811-1877) découvrant par le calcul la planète Neptune illustre bien cette question. Dans le cadre de la mécanique du système solaire, que nous présenterons un peu plus loin, il est apparu que le mouvement d'Uranus ne satisfaisait pas tout à fait à $\vec{F} = m\vec{\gamma}$. Le Verrier attribua cet écart à l'attraction d'une planète inconnue, dont il détermina la masse et la position. Quelques jours après, Neptune était découverte par l'observation.

quelle accélération considérer ?

Nous avons vu précédemment que l'accélération était une notion relative, et notre question sur l'accélération revient à cette autre : dans quel référentiel est vraie notre loi fondamentale? Dans notre description des mouvements des graves, ainsi que dans nos observations des départs de fusées à Cap Canaveral, il s'agissait, bien entendu, du référentiel Terre. Dans notre description des mouvements des satellites, nous avons bien insisté sur le fait qu'il s'agissait du référentiel Terre-Étoiles.

principe de l'inertie

D'ailleurs, une conséquence de la loi fondamentale est la suivante : pas de force, pas d'accélération.



Or, pas d'accélération signifie repos ou mouvement rectiligne uniforme.

Ainsi, un corps qui n'est soumis à aucune force est soit au repos, soit animé d'un mouvement rectiligne uniforme (principe de l'inertie). Mais par rapport à quel référentiel ce principe va-t-il être valable ?

à la recherche du référentiel absolu

Il nous reste donc à chercher le référentiel par rapport auquel est vraie la formule $\vec{F} = m\vec{\gamma}$, où le terme $\vec{F}$ représentera les forces à signification physique concrète et raisonnable décelées expérimentalement. Nous appellerons « absolu » un tel référentiel, et c'est à sa recherche que nous allons procéder.

une mécanique intermédiaire : celle du référentiel « Terre-Étoiles »

Suivant notre idée primitive de ne pas rester trop collés à la Terre, sans cependant nous en éloigner trop considérablement d'un seul coup, nous prendrons comme mécanique de départ celle que nous a amenés à considérer l'étude des satellites artificiels de la Terre. C'est une théorie imparfaite, selon laquelle la Terre a été supposée sphérique et isolée dans l'univers, mais dont le caractère intermédiaire nous permettra de redescendre sur Terre aussi aisément que de remonter vers le système solaire, ou même vers les étoiles. Nous la présenterons schématiquement à l'aide du tableau en bas de la page.

Avant de rectifier cette théorie mécanique un peu sommaire, il nous faut montrer comment elle peut être utilisée dans un autre référentiel. Cela pose le problème du changement de référentiel en dynamique.

que deviennent les forces quand change le référentiel ?

Les données du problème sont celles que nous avons indiquées précédemment à propos de la composition des

Mécanique du référentiel « Terre-Etoiles ».

Référentiel « Terre-Etoiles ».

Loi fondamentale : $\vec{F} = m\vec{\gamma}$

$\vec{\gamma}$ accélération par rapport au référentiel « Terre-Etoiles ». $\vec{F}$: force d'attraction de la gravitation terrestre + toute force décelée ainsi qu'il a été indiqué dans le programme d'utilisation de $F = m\gamma$ (poussée de réacteurs éventuellement mis en marche, résistance de l'air, action d'un champ électromagnétique, etc.).

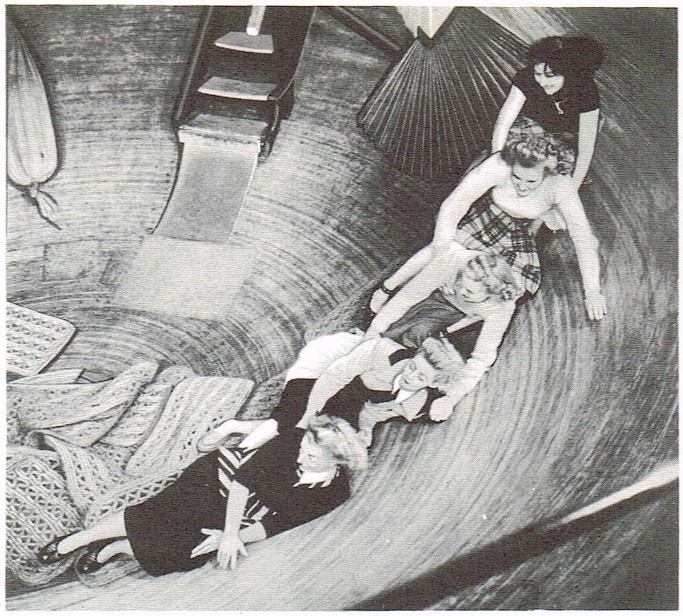
vitesses et des accélérations. Mais nous supposons ici que, par rapport au référentiel absolu R, la loi fondamentale est $\vec{F} = m\vec{\Gamma}_a$, $\vec{\Gamma}_a$ étant l'accélération absolue et $\vec{F}$ la résultante de toutes les forces appliquées au mobile. Si l'on considère un autre référentiel R′, en mouvement connu par rapport à R, cette loi fondamentale s'écrira

$$\vec{F} = m\vec{\Gamma}_r + m\vec{\Gamma}_e + m\vec{\Gamma}_c.$$

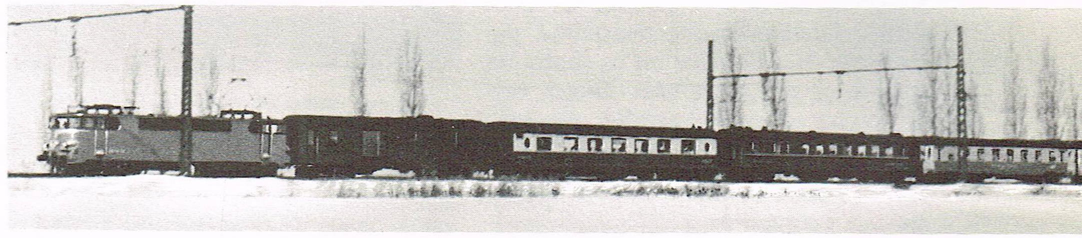
Les observateurs attachés au référentiel R′ mesureront l'accélération relative Γ_r, et écriront donc

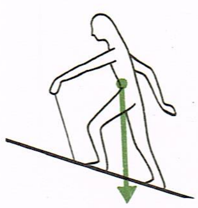
$$\vec{F}\text{-}m\vec{\Gamma}_e - m\vec{\Gamma}_c = m\vec{\Gamma}_r.$$

Ils pourront donc utiliser la loi $\vec{F} = m\vec{\Gamma}_r$, où $\vec{\Gamma}_r$ sera l'accélération relative à leur référentiel, à condition toutefois de faire intervenir dans le premier membre non seulement les forces considérées par les observateurs absolus, mais encore les deux termes $-\ m\vec{\Gamma}_e$ et $-\ m\vec{\Gamma}_c$. On appelle encore ces termes des « forces » et on les qualifie parfois de « fictives », pour les distinguer des forces « réelles » considérées par les observateurs absolus. Ces deux forces, qui sont introduites par le mouvement du nouveau référentiel, s'appellent respectivement la **force d'inertie d'entraînement** et la **force d'inertie complémentaire** (ou « force de Coriolis »). En fait, le mot « fictif » est mal choisi : ces deux forces sont bien réelles; elles se manifestent dans R′ à l'instar des autres forces et y sont de même mesurables.



R est la Terre et R′ une « roue du mystère » que l'on peut voir parmi les attractions foraines, constituée par un cylindre vertical tournant autour de son axe. C'est la force d'inertie d'entraînement qui colle le personnage sur la paroi, où il a l'impression d'être couché comme sur un sol horizontal. Un dynamomètre attaché à la paroi extérieure décèle et mesure cette force.



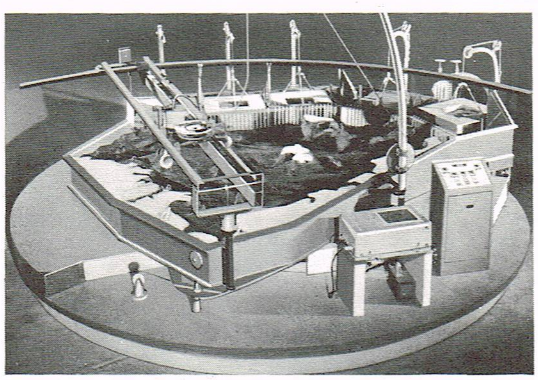


R est la Terre et R′ un train qui accélère en ligne droite. La force d'inertie d'entraînement pousse le voyageur vers l'arrière du train. S'il est assis face à la marche, il est enfoncé dans son dossier. S'il se déplace vers l'avant, cette même force l'oblige à « monter » sur le sol horizontal du wagon. Elle lui permet, en revanche, de « descendre » aisément vers l'arrière du train.



R est la Terre et R′ un navire qui a par rapport à R d'importantes rotations de tangage et de roulis. Le marin qui se déplace sur le navire subit une force d'inertie complémentaire très appréciable, que l'on rend parfois responsable du mal de mer. Il peut annuler cette force — 2 m $\vec{\Omega}\Lambda\vec{V}_r$ en restant étendu sur sa couchette.





Si R est Terre-Étoiles et R′ la Terre, c'est la force d'inertie complémentaire qui fait tourner par rapport à R′ le plan d'oscillation du pendule de Foucault *(v. p. 24)* de telle sorte qu'il soit fixe par rapport à R. C'est elle qui est responsable de la déviation vers l'est des projectiles *(v. p. 24)*. C'est elle qui, dans l'hémisphère Nord, incline les vents vers la droite. C'est pour en rendre compte qu'il a fallu, à Chatou, faire tourner les maquettes sur lesquelles sont faites les études relatives aux usines marémotrices.

que doit faire le Terrien ?

Que doit faire l'habitant de la Terre s'il veut utiliser la mécanique du système Terre - Étoiles, en introduisant l'accélération qu'il mesure, c'est-à-dire l'accélération relative au référentiel Terre ? Nous l'avons déjà vu précédemment, dans le cadre de notre explication par l'accélération, et n'avons ici qu'à le répéter, mais cette fois dans le cadre de l'explication par les forces. Il utilisera l'accélération relative et, dans le terme $\vec{F}$, il introduira le poids du corps (et, s'il s'agit d'expériences très précises, la force d'inertie complémentaire). Le poids est pour lui une donnée expérimentale, mais il pourra lui être agréable de savoir que ce poids est la résultante de l'attraction terrestre et de la force d'inertie d'entraînement, due à la rotation de la Terre autour de la ligne des pôles. Cela lui expliquera la diminution du poids constatée dans un déplacement de l'équateur au pôle et, en utilisant le poids mesuré expérimentalement, il aura l'avantage de se dégager du schéma d'une terre sphérique et homogène. Naturellement, en employant la mécanique du référentiel Terre-Étoiles, il ne peut encore se dégager du schéma de la Terre isolée dans l'univers.

Phot. Berton (« la Vie du rail »), Keystone, X.

que doit faire le cosmonaute ?

Puisque nous avons appris à faire de la mécanique dans n'importe quel référentiel, profitons-en pour faire, toujours à partir de notre mécanique intermédiaire du référentiel Terre-Étoiles, celle du cosmonaute. Nous supposerons que le satellite habité par le cosmonaute ne tourne pas sur lui-même (par rapport au référentiel Terre-Étoiles), ce qui se produit si, au départ, on ne lui a pas communiqué de rotation (toujours par rapport au même référentiel). Il a un mouvement de translation, que nous supposerons circulaire; ses antennes et ses hublots pointent constamment vers des étoiles fixes.

Notre cosmonaute se référera au référentiel Satellite-Étoiles, matérialisé par un triède dont l'origine est au centre du satellite et qui a ses axes dirigés vers des étoiles fixes; le satellite est solidaire de ce référentiel.

Pour un corps de masse m, la loi fondamentale s'écrit, dans le référentiel Satellite-Étoiles,

$$m\vec{A}_T - m\vec{\Gamma}_e = m\vec{\Gamma}_r.$$

$\vec{A}_T$ est l'attraction de la Terre sur la masse-unité située dans le satellite; $\vec{\Gamma}_e$ est l'accélération d'entraînement, c'est-à-dire, puisque le mouvement d'entraînement est une translation, l'accélération (par rapport à Terre-Étoiles) du centre du satellite; $\vec{\Gamma}_r$ est l'accélération de la masse m par rapport à Satellite-Étoiles. Aucune force d'inertie complémentaire ne se fait sentir puisque l'entraînement est de pure translation.

Mais pour le satellite de masse M, en mécanique du référentiel Terre-Étoiles, nous avons : $M\vec{A}_T = M\vec{\Gamma}_e$.

Ainsi, dans le référentiel Satellite-Étoiles, l'attraction terrestre $m\vec{A}_T$ est exactement compensée par la force d'inertie et d'entraînement $- m\vec{\Gamma}_e$. Cette compensation n'est tout à fait rigoureuse que si le corps se trouve au centre des masses du satellite; sinon ce n'est pas exactement le même vecteur $\vec{A}_T$ qui figure dans nos deux formules; nous considérons cette compensation comme exacte à l'intérieur du satellite.

l'apesanteur

Tout corps situé à l'intérieur du satellite n'est soumis à aucune force. Dans le satellite, la loi fondamentale est $\vec{O} = m\vec{\Gamma}_r$.

Les corps passifs y sont en mouvement rectiligne uniforme (jusqu'au choc sur une paroi!) ou au repos, suivant les conditions initiales. C'est l'état d'**apesanteur :** à l'intérieur du satellite, il n'y a aucun champ de forces. Cela ne signifie pas que le cosmonaute n'est pas maître de ses mouvements; il peut exercer des forces par effort musculaire, il peut s'accrocher aux parois et se déplacer le long d'elles, il peut se lancer d'une paroi à l'autre ou lancer des objets.

Cette situation nouvelle pour l'homme est évidemment digne du plus grand intérêt. Nous avons tous été étonnés la première fois que nous avons vu sur une photographie un homme faire l'oiseau à l'intérieur d'un engin en vol. Il est amusant d'imaginer certaines difficultés d'alimentation dans cette situation, qui pose d'importants problèmes biologiques. Il faut, en effet, bien se rendre compte que chaque organe du cosmonaute se trouve dans un champ de force nul, ce qui a suscité de graves inquiétudes. Mais l'adaptation de l'être humain à ces circonstances nouvelles n'a pas été décevante.

Si le satellite est soumis à d'autres forces qu'à l'attraction terrestre, par exemple à la poussée constante $\vec{P}$ d'un réacteur, la loi de son mouvement s'écrira, en mécanique du référentiel Terre-Étoiles, $M\vec{A}_T + \vec{P} = M\vec{\Gamma}_e$, et la loi du mouvement d'un corps inerte à l'intérieur du satellite s'écrira, en mécanique du référentiel Satellite-Étoiles,

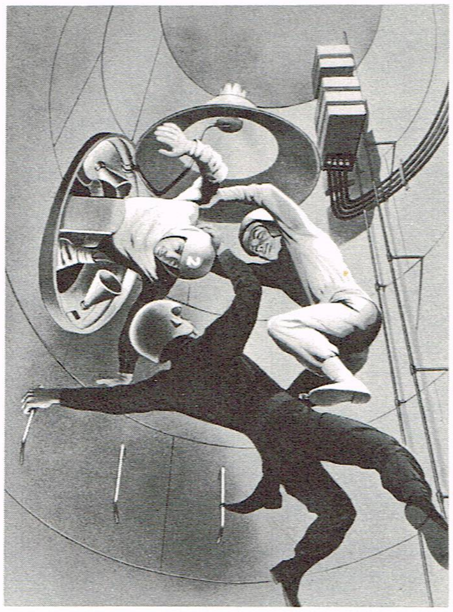
$$- m \frac{\vec{P}}{M} = m\vec{\Gamma}_r.$$

La force du premier membre est analogue à un poids : la poussée P crée une **pesanteur artificielle** à l'intérieur du satellite.

le système solaire

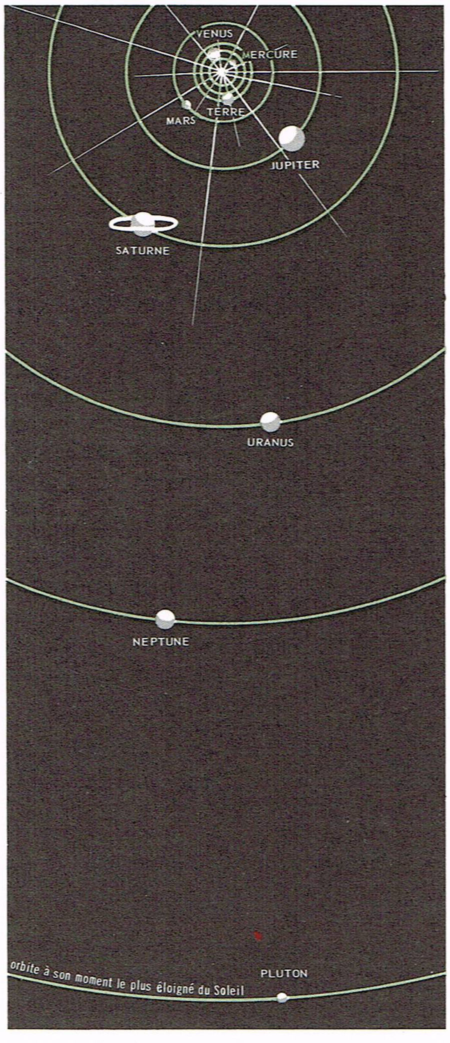
Puisque nous savons que c'est du schéma de la Terre isolée dans l'espace qu'il faut nous dégager, partons dans l'autre sens à la recherche du référentiel absolu et envisageons l'ensemble du système solaire.

Les mouvements des planètes autour du Soleil obéissent aux mêmes lois (de Kepler) que les mouvements des satellites artificiels autour de la Terre. Reprenant alors ce qui a été dit à propos



Astronautes en état d'apesanteur.

Dessin de Walt Disney.



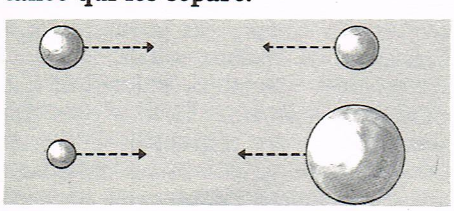
Plan du système solaire.

de ces derniers mouvements, nous pouvons expliquer les mouvements des planètes. Dans le référentiel constitué par le centre du Soleil et les directions des étoiles fixes, par la gravitation solaire : le Soleil attire les planètes avec une force inversement proportionnelle au carré de la distance de la planète au Soleil.

la gravitation universelle

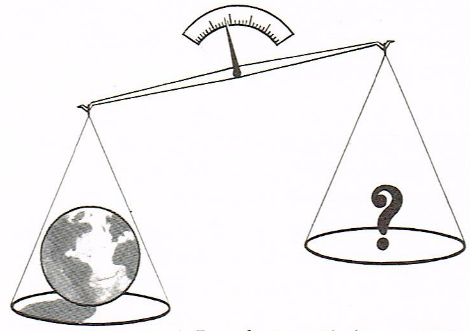
Le phénomène de gravitation a été rencontré à diverses reprises, et il convient à présent de faire une synthèse. L'expérience de Cavendish, qui est une expérience de statique, aurait pu être utilisée pour mettre en évidence la validité du principe de l'égalité de l'action et de la réaction pour les forces gravitationnelles. Le corps B suspendu au pendule de torsion était attiré par le corps A maintenu immobile. En fait, A est attiré par le corps B avec une force égale à l'attraction exercée par A sur B.

Il convient, d'autre part, d'étendre à la dynamique le principe de l'égalité de l'action et de la réaction. C'est ainsi seulement que ce principe peut avoir une portée générale, tout mouvement étant relatif. Le phénomène de gravitation est donc régi, indépendamment de la position dans l'espace, de la configuration des masses, de l'état de mouvement de ces masses et de l'époque considérée, par la loi de la gravitation universelle, énoncée par Newton : **Deux corps de l'univers s'attirent réciproquement en exerçant l'un sur l'autre deux forces opposées sur la droite qui les joint, proportionnelles aux masses de ces corps et inversement proportionnelles au carré de la distance qui les sépare.**

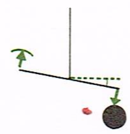


peser la Terre !

Mais nous avons vu le lien qui existe entre l'attraction terrestre et la pesanteur. Si l'on néglige la force d'inertie d'entraînement et si l'on se place à l'un des pôles, le poids *mg* d'un corps est la force avec laquelle il est attiré par la Terre, c'est-à-dire, d'après la loi de Newton, $k\frac{Mm}{R^2}$ (M et R : masse et rayon de la Terre). On démontre, en effet,



peser la Terre: 6 suivi de 21 zéros, en tonnes



$k = 6{,}67 \cdot 10^{-8}$

$g = 981$

$R = 6370 \cdot 10^5$

$k\frac{M}{R^2} = g$

$M = 6 \cdot 10^{21}$ tonnes

— en attribuant à la Terre une forme sphérique et une structure homogène —, que l'attraction qu'elle produit est égale à celle qu'elle exercerait si toute sa masse était concentrée en son centre. Comme *g* est connu, ainsi que R, et que l'expérience de Cavendish nous a fourni la valeur de la constante de la gravitation *k*, nous pouvons dire que cette célèbre expérience permet de peser la Terre (déterminer sa masse).

Connaissant la masse et le rayon de la Terre, on en déduit la densité moyenne

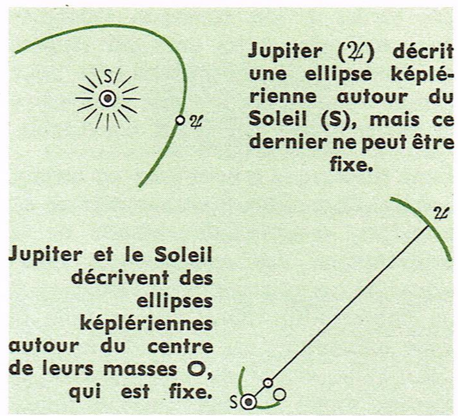
$$\rho = \frac{M}{\frac{4}{3}\pi R^3} = \frac{6 \cdot 10^{21}}{\frac{4}{3}\pi (6\,370\,000)^3}$$

= 5,5 tonnes par mètre cube.

Remarquons qu'une telle expérience nous renseigne sur ce qui se passe à l'intérieur même de la Terre et que, pour être indirect, ce procédé d'investigation (un des seuls possibles à notre disposition, avec la gravimétrie et l'étude des secousses telluriques et de leur propagation) n'en est pas moins précieux.

la mécanique universelle

Puisque nous admettons qu'une planète qui est attirée par le Soleil l'attire elle-même avec autant de force, nous ne pouvons plus considérer le Soleil comme fixe; cela serait en contradiction avec la loi fondamentale qui nous assure qu'à une force non nulle correspond une accélération non nulle. Pour un système solaire qui serait réduit au Soleil et à une planète (par exemple la plus grosse : Jupiter), on démontre que pour avoir des résultats en accord avec les observations, c'est le centre des masses du Soleil et de Jupiter qui doit être considéré comme fixe.



De ces deux conceptions, qui rendent aussi bien compte l'une que l'autre des observations, c'est la seconde qui doit être retenue à cause de la loi fondamentale.

Pour le système solaire tel qu'il est constitué en réalité, c'est le centre des masses de tout le système qu'il convient de considérer comme fixe, et nous trouvons enfin le référentiel convenable pour la mécanique qui régit les mouvements des corps à l'intérieur du système solaire : c'est le **référentiel de Copernic,** matérialisé par un trièdre ayant son origine au centre des masses du système solaire et ses axes dirigés vers des étoiles fixes.

LA MÉCANIQUE UNIVERSELLE
Référentiel : celui de Copernic.
Loi fondamentale : F = $m\Gamma$.
Γ : accélération par rapport au référentiel de Copernic.
F : toutes les forces « réelles ».

Présentons à ce sujet quelques remarques.

Cette mécanique « universelle » est valable pour l'univers réduit au système solaire. On y a admis la fixité des étoiles, nous contentant d'assimiler l'univers extérieur au système solaire à une pure figure de géométrie. En fait, des mouvements des étoiles ont été décelés, et la loi de la gravitation de Newton a été étendue à tous les astres de l'univers. Le Soleil est une petite étoile de notre Galaxie, qui n'est elle-même qu'un amas stellaire parmi ceux qui constituent l'univers. La rotation d'ensemble de notre Galaxie, les mouvements propres

des étoiles à son intérieur, l'éloignement des galaxies les unes par rapport aux autres sont des phénomènes pour l'étude desquels il faudrait poursuivre la recherche du référentiel convenable.

Cette mécanique a pour base un certain nombre de notions fondamentales et de principes, et sa validité résulte de la confrontation des conséquences théoriques de ces principes avec l'expérience ou l'observation. Nous tenons enfin la vraie mécanique, car on peut dire que, dans l'état actuel de nos moyens de mesure, toutes les expériences et toutes les observations sont conformes aux conséquences des principes de cette mécanique (1).

Enfin, cette mécanique, qui est la bonne, va pouvoir être utilisée dans n'importe quel référentiel, puisque le problème du changement de référentiel en dynamique a été présenté.

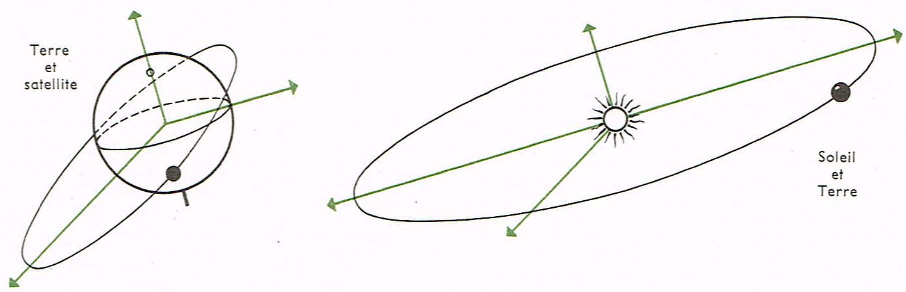
référentiels galiléens

Les deux forces fictives à considérer sont nulles si le nouveau référentiel a un mouvement de translation rectiligne uniforme par rapport au référentiel de Copernic. De tels référentiels sont appelés **galiléens.** La mécanique universelle est aussi bien valable avec la même loi et les mêmes conditions d'utilisation dans tout référentiel galiléen que dans le référentiel même de Copernic. Cela est quelque peu surprenant si l'on pense que la vitesse du référentiel galiléen par rapport au référentiel de Copernic peut être « aussi grande que l'on veut ». Aucune expérience de pure mécanique n'est susceptible de mettre en évidence un mouvement de translation rectiligne uniforme. La voie s'ouvre vers un autre chapitre de cet ouvrage, où il sera fait appel aux phénomènes électromagnétiques.

référentiel « Terre-Étoiles » et référentiel « Terre »

Passons à présent du référentiel de Copernic (ou d'un référentiel galiléen) R au référentiel Terre-Étoiles R′. R′ est en mouvement de translation par rapport à R : il n'y a pas d'accélération complémentaire.

Si le Soleil et la Terre constituaient à eux seuls le système solaire, la situation du référentiel Terre-Étoiles par rapport au référentiel de Copernic serait exactement la même que la situation étudiée antérieurement du référentiel Satellite-Étoiles par rapport au référentiel Terre-Étoiles. Ce que nous avons développé à ce sujet et qui était relatif à l'état d'apesanteur du cosmonaute peut se transposer dans notre problème actuel. Les observateurs du référentiel Terre-Étoiles se trouvent en état d'absence de gravitation solaire, la force attractive du Soleil étant compensée par la force d'inertie d'entraînement.



Notre schéma de la Terre isolée dans l'univers n'était donc pas si mauvais. S'il n'était pas rigoureusement exact, cela tenait à ce que la compensation entre l'attraction solaire et la force d'inertie complémentaire n'est rigoureuse qu'au centre de la Terre et que nous avions négligé la présence des autres astres du système solaire, c'est-à-dire par ordre d'importance de leurs actions, la Lune, puis les autres planètes.

Il s'introduit donc en toute rigueur une correction excessivement faible dans ce que nous avons appelé la mécanique du référentiel Terre-Étoiles. Cette correction est variable avec l'époque considérée, puisqu'elle dépend de la position dans le ciel de la Lune, du Soleil et des planètes.

Cette correction se répercute sans modification dans la mécanique du référentiel Terre. Mais comme elle intervient naturellement dans l'expérience fondamentale de la statique qui définit le poids d'un corps, elle entre de ce fait dans la définition rigoureuse du poids.

Il peut être agréable de savoir que dans la définition rigoureusement élaborée de la notion du poids entrent :

l'attraction des masses terrestres, qui sont loin d'être uniformément réparties dans une Terre qui n'est pas sphérique;

la force d'inertie d'entraînement, due au mouvement de rotation de la Terre autour de la ligne de ses pôles;

une très légère différence entre l'attraction de la Lune, du Soleil et des planètes et la force d'inertie d'entraînement due au mouvement de translation de la Terre sur son orbite autour du Soleil.

Mais cela n'est pas utile à beaucoup de personnes. Ce n'est que dans l'étude des marées qu'intervient, pour les phénomènes ayant lieu à la surface de la Terre, le troisième élément du poids figurant ci-dessus. Le poids est, heureusement pour nous, une donnée expérimentale globale que la plupart du temps il n'y a pas à couper en trois ou en quatre! Mais pouvait-on imaginer *a priori* combien est complexe l'interprétation du fait qu'un corps suspendu à un dynamomètre produit un allongement du ressort?

Nous sommes donc arrivés au terme de notre voyage à la recherche du référentiel absolu par rapport auquel il convient d'envisager les notions, les principes et la loi fondamentale de la mécanique. Nous étant limités volontairement à vouloir rendre compte des mouvements à l'intérieur du système solaire, la mécanique est la mécanique universelle que nous avons présentée. Celle-ci peut être utilisée dans des référentiels différents et avec des approximations différentes suivant les différentes catégories de mouvement que l'on se propose d'étudier. Non seulement, d'ailleurs, le cadre et l'approximation à choisir dépendront de la classe de mouvements étudiée, mais encore en dépendront les méthodes de traitement des équations différentielles qui servent à la prévision du mouvement. Nous ne pouvons insister sur ces méthodes, qui sont bien souvent des méthodes d'approximations successives, bien adaptées au caractère expérimental de la science mécanique.

(1) Trois désaccords subsistent cependant, dont le plus important est relatif à l'avance du périhélie de Mercure. Aujourd'hui, la mécanique classique de Newton ne doit plus être considérée que comme une première approximation d'une théorie beaucoup plus évoluée : la théorie relativiste de la gravitation, d'Einstein. En dehors du fait que cette théorie est susceptible de rendre compte du désaccord signalé, elle va beaucoup plus loin dans l'explication du phénomène de gravitation. Mais pour tous les besoins du mécanicien, la théorie newtonienne apparaît comme suffisante; on lui ajoute simplement, dans les seuls cas où cela est nécessaire, une correction relativiste. (Voir chapitre RELATIVITÉ.)

les équations du mouvement

Qui n'a entendu parler d'énergie, de travail, d'impulsion, de force vive, de quantité de mouvement — tous mots passés dans le langage le plus courant sans que l'on en sache bien la signification exacte? Tout au plus est-il évident qu'ils ont rapport avec des efforts à fournir, des forces à exercer plus ou moins brutalement.

Il semble donc naturel de leur donner un sens mécanique, de les rattacher à la relation fondamentale dont la dynamique nous a armés, et nous verrons que, grâce à eux, nous pourrons aller plus avant dans la compréhension des phénomènes de la mécanique.

Nous avons assisté à l'élaboration de la loi $\vec{F} = m\vec{\gamma}$, nous avons vu quelle signification il convenait de lui attribuer, dans quel cadre la situer. Mais nous n'avons guère vu comment l'utiliser commodément.

Progressivement, à partir du XVIIIe siècle, des notions mécaniques se sont dégagées au prix de discussions souvent âpres et laborieuses. Aux concepts de force, d'accélération et de masse, se sont ajoutés ceux de quantité de mouvement, de moment cinétique et d'énergie cinétique (la **cinétique** étant la science du mouvement des masses).

Ces éléments cinétiques ont une signification physique précise et importante en ce qui concerne l'état de mouvement des masses. Ils ne font intervenir que la masse et la vitesse du mobile. Les deux premiers sont des quantités vectorielles et le troisième une quantité scalaire (numérique).

L'évolution au cours du temps de ces éléments cinétiques est commandée par la loi fondamentale; les lois de cette évolution ne seront donc que des conséquences, ou d'autres aspects, de la formule fondamentale.

quantité de mouvement

Transportons-nous de nouveau à Cap Canaveral, lors de lancements de fusées. Pour leur communiquer une certaine vitesse, il faut exercer sur elles une force, mais leur masse intervient de façon essentielle : il est naturellement plus difficile de communiquer la première vitesse spatiale au satellite « Score » de 4 000 kg qu'à « Explorer I » de 14 kg; comme il est plus difficile de communiquer la même vitesse initiale au poids réglementaire des compétitions athlétiques qu'à un petit caillou.

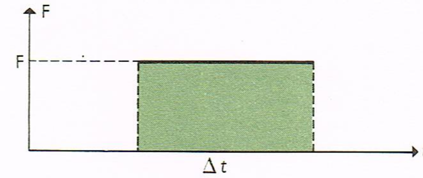
Le problème n'est pas de communiquer une certaine vitesse, il est de communiquer à une certaine masse une certaine vitesse. Les deux quantités m et $\vec{V}$ doivent intervenir au même titre et c'est en fait le produit $m\vec{V}$ de ces deux quantités qu'il convient de considérer. Ce produit de la masse par la vitesse s'appelle la **quantité de mouvement** du corps; c'est une grandeur vectorielle permettant d'estimer le mouvement de la masse.

Son évolution au cours du temps est donnée directement par la loi fondamentale, que l'on peut écrire sous la forme $\vec{F} = \frac{d}{dt}(m\vec{V})$, puisque la masse est, en mécanique classique, une quantité invariable. Nous avons ainsi un nouvel énoncé de la loi fondamentale : **la dérivée de la quantité de mouvement est égale à la résultante des forces appliquées.**

impulsion

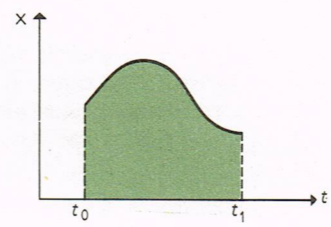
La quantité de mouvement, élément cinétique, est en relation avec un élément dynamique, l'impulsion. Si l'on ne s'intéresse qu'au résultat de l'opération de lancement, le problème est le suivant : faire passer une masse m de l'état de repos (quantité de mouvement nulle) à un état de vitesse $\vec{V}$ (quantité de mouvement $m\vec{V}$).

Si c'est une force constante qui a opéré pendant un intervalle de temps Δt, la quantité de mouvement communiquée est égale au produit de la force par son temps d'action : $\vec{F}\Delta t = m\vec{V}$. En effet, si la force est constante, les deux quantités $d(m\vec{V})$ et dt intervenant dans la formule $\vec{F} = \frac{d(m\vec{V})}{dt}$ restent tout le temps proportionnelles. Leurs variations finales $\Delta(m\vec{V})$ et Δt sont dans le même rapport : $\Delta(m\vec{V}) = \vec{F}\Delta t$. La quantité $\vec{F}\Delta t$ s'appelle l'*impulsion* communiquée au corps par la force $\vec{F}$ agissant pendant le temps Δt.



L'aire colorée représente la grandeur de l'impulsion.

On peut la définir d'une façon plus générale : lorsque la force n'est pas constante, on découpe le temps pendant lequel elle agit en petits intervalles pendant lesquels elle peut être considérée comme constante et l'on fait le total des variations correspondantes. L'impulsion est une quantité vectorielle $\vec{\mathscr{F}}$ dont on calcule les trois composantes $\mathscr{X}$, $\mathscr{Y}$, $\mathscr{Z}$, en intégrant les trois composantes X, Y, Z, de la force $\vec{F}$.



L'aire colorée représente la première composante $\mathscr{X}$ de l'impulsion d'une force variable entre les instants t_0 et t_1; c'est l'intégrale de X entre t_0 et t_1.

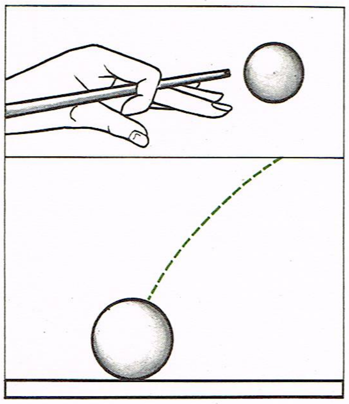
On écrit $\mathscr{X} = \int_{t_0}^{t_1} X dt$, et l'on a deux formules analogues pour $\mathscr{Y}$ et $\mathscr{Z}$.

On notera $\vec{\mathscr{F}} = \int_{t_0}^{t_1} \vec{F} dt$, et l'on dira que l'impulsion $\vec{\mathscr{F}}$ est l'intégrale du vecteur $\vec{F}$ entre t_0 et t_1.

La loi fondamentale de la dynamique peut encore s'énoncer ainsi : l'impulsion communiquée au corps pendant un intervalle de temps quelconque est égale à la variation de la quantité de mouvement. Il s'agit, bien entendu, de l'impulsion totale, c'est-à-dire de l'impulsion de la résultante de toutes les forces appliquées.

cas du choc

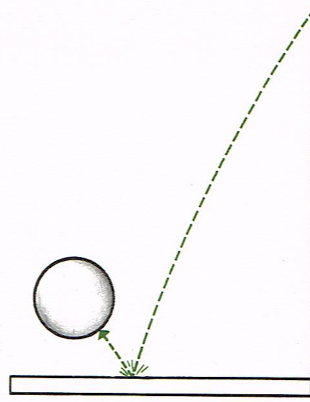
Il y a des cas où une modification très nette de la quantité de mouvement est obtenue d'une façon quasi intantanée par une force très grande, agissant pendant un temps très court. Ce sont les phénomènes de choc. Ils sont régis, en considérant l'impulsion instantanée $\vec{\mathscr{F}}$, par la loi $m\vec{V}_1 - m\vec{V}_0 = \vec{\mathscr{F}}$, $m\vec{V}_0$ et $m\vec{V}_1$ étant les quantités de mouvement avant et



Bille de billard frappée par la queue.

Choc d'une bille d'acier sur un sol en ciment : la quantité de mouvement de la bille se trouve renversée instantanément.

Choc d'une bille de plomb sur un sol de ciment : la quantité de mouvement de la bille s'annule instantanément.

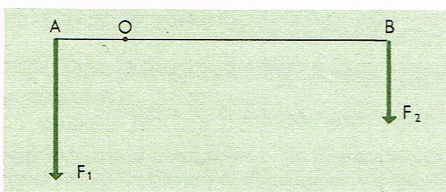


Le résultat du choc dépend essentiellement de la matière dont sont constitués les corps qui se heurtent.

après le choc. Cette loi est plus simple que la loi fondamentale de la dynamique : elle ne fait pas intervenir l'accélération et fournit sans intégration la vitesse du corps après le choc.

moment cinétique

Avant de présenter le deuxième élément cinétique, le moment cinétique, il nous faut connaître l'élément dynamique correspondant : le moment d'une force.
Considérons l'équilibre d'un levier sous l'action de deux forces parallèles $\vec{F}_1$ et $\vec{F}_2$. La condition d'équilibre, dans la disposition de l'appui O et des points



d'application A et B des forces représentées sur la figure, s'énonce

$$OA.\ F_1 = OB.\ F_2.$$

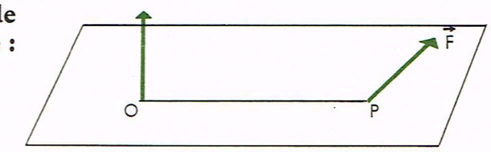
Une force déterminée a d'autant plus d'importance que son point d'appui est plus loin de O. Cette importance d'une force par rapport à un point déterminé O se mesure par le moment de cette force par rapport à O, qui est le produit de la force par le bras de levier : le moment de $\vec{F}_1$ par rapport à O est le produit de son intensité par le bras de levier OA.

D'une façon plus précise, le moment d'une force $\vec{F}$, appliquée en un point P, par rapport à O est le produit vectoriel $\overrightarrow{OP} \wedge \vec{F}$; c'est donc un vecteur d'origine O perpendiculaire au plan défini par la force et le point O.

En introduisant de manière analogue le moment cinétique $\overrightarrow{O\sigma}$ par rapport à un point fixe O comme produit vectoriel de $\overrightarrow{OP}$ et de la quantité de mouvement, on démontre que la **dérivée du moment cinétique d'un mobile par rapport à un point fixe O est égale au moment (par rapport à ce point) de la force totale qui lui est appliquée :**

$$\overrightarrow{OP} \wedge \vec{F} = \frac{d}{dt}(\overrightarrow{O\sigma}).$$

$$\overrightarrow{OP} \wedge \vec{F} = \mathcal{M}_0^t(\vec{F})$$



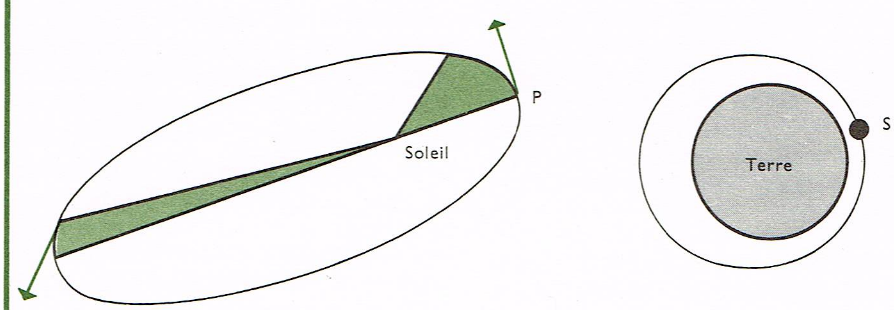
$$\mathcal{M}_0^t(\vec{F}) = \overrightarrow{OP} \wedge \vec{F} = \overrightarrow{OP} \wedge \frac{d}{dt}(m\vec{V}) = \frac{d}{dt}(\overrightarrow{OP} \wedge m\vec{V}) = \frac{d}{dt}\overrightarrow{O\sigma}$$

puisque (règle de dérivation d'un produit)

$$\frac{d}{dt}(\overrightarrow{OP} \wedge m\vec{V}) = \frac{d}{dt}(\overrightarrow{OP}) \wedge m\vec{V} + \overrightarrow{OP} \wedge \frac{d}{dt}(m\vec{V})$$

et que le premier produit vectoriel du second membre est nul (vecteurs parallèles).

loi des aires et champ de forces newtonien

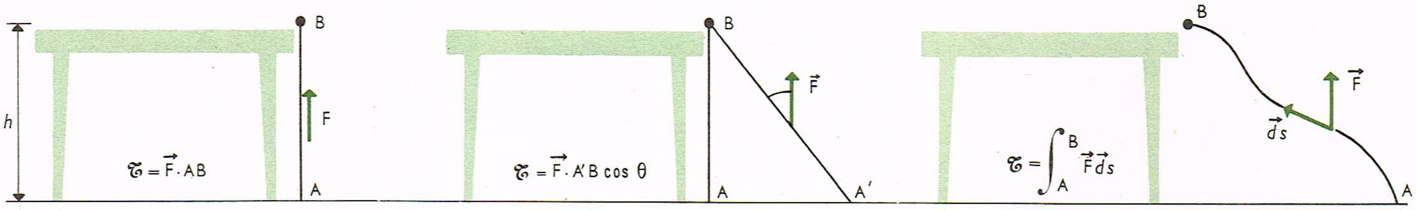


Pour tout champ de forces central, et en particulier pour le mouvement képlérien, une planète P autour du Soleil, ou un satellite S autour de la Terre, le moment de la force par rapport au centre d'attraction est nul. Le moment cinétique par rapport à ce centre est constant : le rayon vecteur balaie des aires égales en des temps égaux. On retrouve la loi des aires : plus le mobile est éloigné du centre, plus il va lentement.

travail d'une force

Ici encore, avant de présenter le troisième élément cinétique, l'énergie cinétique, il nous faut présenter d'abord l'élément dynamique correspondant, le travail d'une force.

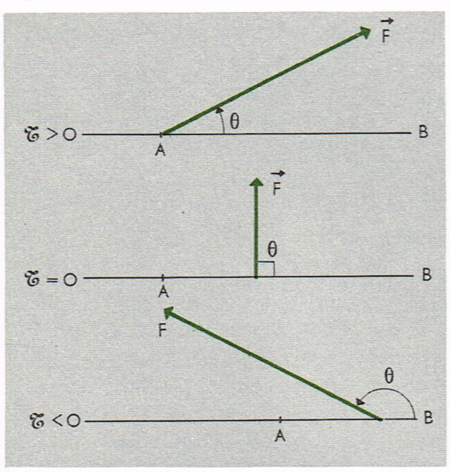
Si je ramasse un objet par terre et que je le place sur une table, j'exerce sur lui une certaine force, et le point d'application de cette force subit un certain déplacement. Je dis que j'ai fourni un certain travail, et ce travail est d'autant plus grand que la force est grande et



que le déplacement vertical est grand. On exprime ce travail par le produit de la force par le déplacement.

Mais le travail dépend des directions relatives de la force exercée et du déplacement : déplacer un objet horizontalement sur un sol bien lisse est aisé. Et l'on paie le marchand de charbon au nombre de sacs transportés et au nombre d'étages gravis, non au nombre de mètres parcourus. Ce qui coûte, c'est de déplacer une force parallèlement à sa direction, et non perpendiculairement. Plus exactement, le travail est le produit scalaire de la force $\vec{F}$ par le déplacement, c'est-à-dire que pour un déplacement rectiligne A′ B d'une force constante, il vaut : $\mathcal{T} = \vec{F}.\,A'B \cos\theta = \vec{F}.\,\overrightarrow{A'B}$ (θ étant l'angle entre la direction de la force et celle du déplacement).

Le travail est une quantité algébrique positive, négative ou nulle, selon que l'angle que font $\vec{F}$ et le déplacement est aigu, obtus ou droit.



Pour une force et un déplacement quelconques, on additionne les travaux qui correspondent aux petits déplacements ds, ce qui conduit à l'expression (intégrale) $\mathcal{T} = \int_A^B \vec{F}\,\vec{ds}$.

énergie cinétique

Il est bon de revenir maintenant au problème de lancement d'une fusée sous l'action d'une poussée (constante) F pour nous intéresser, non plus au temps pendant lequel cette force agit, mais surtout à l'altitude qu'elle permet à la fusée d'atteindre. La poussée, parallèle à la trajectoire, fournit le travail

$$\mathcal{T}(h) = \vec{F}.\,h.$$

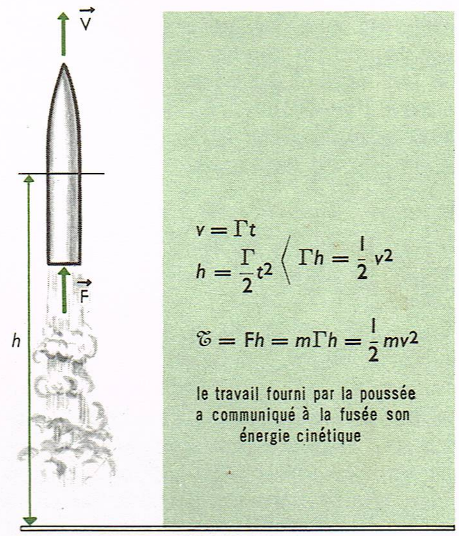
A quoi donc ce travail a-t-il servi, et comment peut-on interpréter son action ? Un calcul simple montre que

$$\mathcal{T}(h) = \vec{F}h = \frac{1}{2}\,m\,v^2.$$

Nous voici donc en présence d'un nouvel élément cinétique important : l'**énergie cinétique,** demi-produit de la masse par le carré de la vitesse (ou encore demi-*force vive*). L'importance de cet élément cinétique est due au fait qu'il est en relation directe avec le travail : le travail (l'effort fourni) ne se perd pas; il se transforme en énergie (cinétique). D'une façon générale, dans tout mouvement, **le travail de la force totale appliquée au mobile pour un certain déplacement du mobile sur sa trajectoire est égal à la variation que subit l'énergie cinétique du mobile dans ce déplacement.** Si ce travail est positif, il y a un effort effectivement fourni au corps, et son énergie cinétique augmente. Si le travail est négatif, c'est le corps qui fournit du travail au détriment de son énergie cinétique : $\mathcal{T} = \Delta\left(\frac{1}{2}mv^2\right)$.

travail et énergie

Le travail fourni par une force peut provenir de différentes sources : les êtres vivants sont capables de fournir du travail; ils sont aussi capables d'en produire à l'aide de machines. Ce travail, qui est susceptible de se transformer en énergie, provient lui-même d'une énergie, et c'est une sorte d'énergie, musculaire, calorifique, électrique, chimique, etc., que l'homme transforme en travail. Il cherche donc à connaître et à utiliser toutes les sources d'énergie. Nous n'insisterons pas sur les différentes sortes d'énergie, cela nous ferait par trop sortir du cadre de la mécanique classique, et ne parlerons ici que d'énergie mécanique. Nous connaissons à présent l'énergie cinétique, susceptible d'être créée par un travail, ou de se transformer en travail. Il existe une autre sorte d'énergie



mécanique : l'énergie potentielle. Avant de la présenter, il nous faut nous familiariser avec la notion d'un champ de forces.

champ de forces et potentiel

Une force appliquée à un corps en mouvement peut dépendre de la position du corps, de sa vitesse et de l'époque considérée. Limitons-nous au cas où elle ne dépend que de la position du corps : nous disons alors que nous avons affaire à un champ de forces.

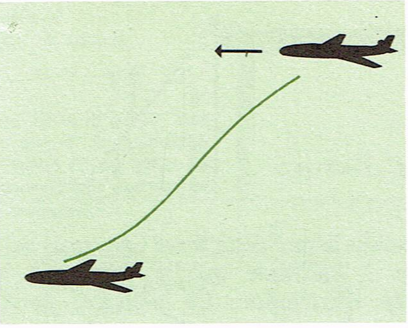
Nous connaissons le champ newtonien de gravitation terrestre; il exerce une force attractive centrale, de grandeur $\frac{\mu}{r^2}$, sur toute masse-unité distante de r du centre de la Terre.

Ce que nous voulons ici, c'est exprimer simplement le travail correspondant au déplacement de cette masse. La force n'étant pas constante, il faut recourir au calcul intégral, qui, seul, est rigoureux. Mais point n'est besoin d'être éminent mathématicien pour deviner la réponse : seul intervient le produit de la force par la variation du rayon :

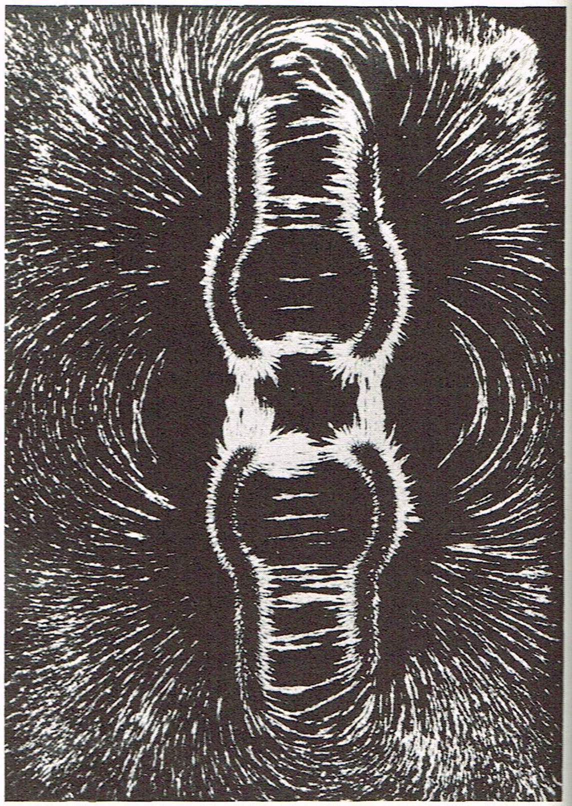
$$-\frac{\mu}{r^2}\,\Delta r = \Delta\left(\frac{\mu}{r}\right).$$

La théorie de l'intégration confirme rigoureusement que $\mathcal{T} = \Delta\left(\frac{\mu}{r}\right)$.

En appelant **potentiel** la quantité $-\frac{\mu}{r}$ et **chute de potentiel** l'opposé de sa variation, on a $-\Delta\left(-\frac{\mu}{r}\right) = \Delta\left(\frac{\mu}{r}\right)$, et l'on peut énoncer : **le travail de force d'un champ newtonien lorsqu'on déplace son point d'application de P_0 à P_1 ne dépend que des points d'arrivée et de départ** (et non du chemin parcouru); **il est égal à la chute de potentiel entre l'arrivée et le départ.** On dit encore, plus simplement, que le champ **dérive d'un potentiel.**



L'avion tombe dans un trou d'air de 800 m à 720 m d'altitude; sa chute équivaut à 800 — 720 = 80 m, alors que la variation de son altitude est de 720 — 800 = — 80 m.



Photographie d'un champ magnétique.

champ dérivant d'un potentiel : lignes de forces

Dans le cas d'un champ de forces général, les lignes qui sont en tout point tangentes à la force issue de ce point sont appelées **lignes de forces** du champ. Ce sont elles que matérialisent les grains de limaille de fer placés dans le champ magnétique d'un aimant.

Les lignes de forces existent dans n'importe quel champ. Mais la propriété du travail d'être indépendant du chemin ou, ce qui revient au même, l'existence d'un potentiel, n'a pas lieu pour tous les champs de forces.

Si le champ dérive d'un potentiel, la connaissance de ce potentiel suffit pour connaître toutes les propriétés du champ. Bornons-nous au cas d'un champ plan, région plane en chaque point de laquelle est définie une force située dans ce plan.

Les lignes le long desquelles il n'y a pas de chute de potentiel, c'est-à-dire où le potentiel V demeure constant, s'appellent les **lignes de niveau** du champ. Si l'on se déplace sur une telle ligne, le potentiel ne varie pas, la force du champ ne travaille pas; cela signifie que la force fait un angle droit avec cette ligne de niveau.

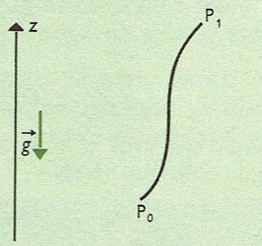
Les lignes de niveau, régulièrement échelonnées en potentiel permettent d'obtenir une représentation graphique du champ de forces, exactement comme on représente la topographie d'un terrain sur une carte à l'aide de lignes de niveau régulièrement échelonnées en altitude.

l'énergie potentielle

Considérons à présent le mouvement d'un mobile dans un champ de forces dérivant d'un potentiel, et supposons que ce mobile n'est soumis qu'à la

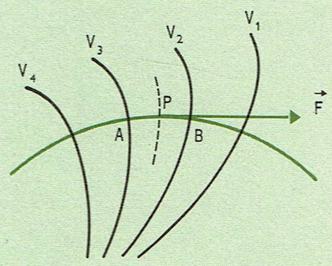
Champ constant.

Champ de la pesanteur dans une région pas trop vaste à la surface de la Terre.
Le travail de la force g du champ dans le passage du point P_0 d'altitude z_0 au point P_1 d'altitude z_1 est $g(z_0 - z_1)$. Il y a un potentiel qui est gz.



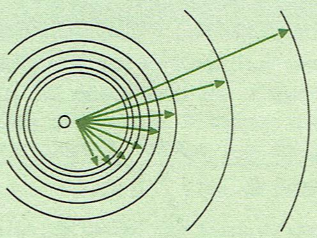
Cas général.

Faisons le choix d'un certain saut de potentiel h et traçons les lignes de niveau $V_1 = h$, $V_2 = 2h$, $V_3 = 3h$... En un point P, la force est perpendiculaire à la ligne de niveau qui passe par P. La force est dirigée vers les potentiels décroissants puisque à une chute de potentiel correspond un travail positif de la force. Elle est en grandeur inversement proportionnelle à l'écartement AB = e des lignes de niveau, puisque, pour leur déplacement AB, le travail est F.e = h. La force est intense là où les lignes de niveau sont serrées, faible là où elles sont écartées.



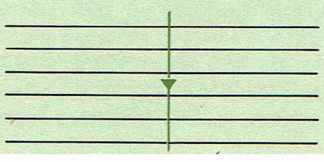
Champ newtonien.

Dans un plan passant par le centre des forces O, les lignes de forces sont les rayons issus de O. Les lignes de niveau, les cercles de centre O.



Champ de la pesanteur.

Dans un plan vertical, les lignes de forces sont les verticales, les lignes de niveau les horizontales.



seule force de ce champ. Nous pouvons alors appliquer à ce mobile la relation trouvée précédemment entre le travail de la force totale et l'énergie cinétique.

Raisonnons sur les mouvements des satellites artificiels dans le champ de gravitation terrestre. Dans un mouvement circulaire, qui est uniforme, l'énergie cinétique $\frac{1}{2} mv^2$ ne varie pas, et la force d'attraction ne travaille pas puisqu'elle est toujours perpendiculaire au déplacement. Mais, dans un mouvement elliptique, la grandeur de la vitesse varie, donc aussi l'énergie cinétique. La force de gravitation travaille, et son travail est donné par la chute du potentiel newtonien $-\frac{\mu}{r}$. Pour un satellite de masse m, la force de gravitation est $\frac{m\mu}{r^2}$, et c'est la chute de la quantité $-\frac{m\mu}{r}$ qui donne le travail de la force de gravitation. Cette quantité, dont la chute représente un travail, est donc une énergie, c'est ce qu'on appelle l' « énergie potentielle » du corps. L'énergie potentielle n'est définie qu'à une constante additive près ; seules les variations d'énergie potentielle ont un sens physique. Du seul fait qu'un corps se trouve dans un champ de forces, il possède une certaine forme d'énergie : l'**énergie potentielle.** Cela veut dire que le corps placé dans le champ a, de ce fait même, une certaine possibilité inhérente. Cette possibilité peut se manifester soit en se transformant en énergie cinétique, si le corps se met en mouvement, soit en fournissant du travail, par exemple celui qui est nécessaire à l'allongement d'un ressort ou à la déformation d'un appui, si le corps est immobilisé.

Pour les mouvements qui ont lieu dans un champ de forces dérivant d'un potentiel, il y a donc une relation très simple entre l'énergie cinétique et l'énergie potentielle : la variation de l'énergie cinétique est égale à la chute d'énergie potentielle. En tenant compte des sens qui ont été précisés pour les mots « variation » et « chute », cela indique que toute augmentation d'une de ces énergies est compensée par la diminution de l'autre. On définit la somme de ces deux énergies comme l'énergie totale du corps. Ainsi, dans tout mouvement dans un champ de forces dérivant d'un potentiel, l'énergie totale du corps demeure invariable.

Arrêtons-nous à cet aspect du théorème de l'énergie.

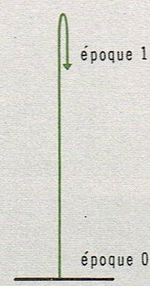
à propos du théorème de l'énergie

Le théorème de l'énergie dans le cas qui vient d'être considéré donne une relation entre la vitesse du corps et sa position. Alors que la loi fondamentale fait intervenir la dérivée seconde de la position du mobile, cette relation ne fait intervenir que la dérivée première.

On dit qu'une telle relation constitue une **intégrale première** du mouvement.

La recherche et l'utilisation des intégrales premières d'un mouvement sont d'une importance considérable dans les problèmes de mécanique.

Le théorème de l'énergie indique qu'une certaine quantité reste invariable au cours du mouvement, que l'énergie totale se conserve. La recherche et l'étude des quantités qui se conservent au cours du mouvement constituent un important chapitre de la mécanique.



Quelle altitude maximale atteint un projectile lancé verticalement avec une vitesse v_0 (dans le vide)?

$$\frac{1}{2} m v_1^2 + mg z_1 = \frac{1}{2} m v_0^2 + mg z_0 ; \text{ avec } z_0 = 0 \text{ et } v_1 = 0,$$

$$z_1 = \frac{v_0^2}{2g}.$$

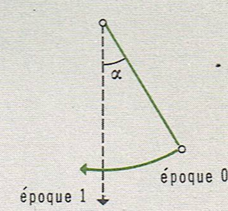
Vitesse de la masse pendulaire en son point le plus bas, lorsque le pendule a été écarté d'un angle α de sa position d'équilibre et lâché sans vitesse.

La force totale est constituée par le poids du corps et la tension du fil. Cette tension ne travaille pas, car elle est toujours perpendiculaire au déplacement.

$$\frac{1}{2} m v_1^2 + mg z_1 = \frac{1}{2} m v_0^2 + mg z_0 ;$$

$$\text{avec } z_1 = 0, z_0 = l(1 - \cos\alpha) \text{ et } v_0 = 0,$$

$$v_1 = \sqrt{2 gl (1 - \cos\alpha)}.$$



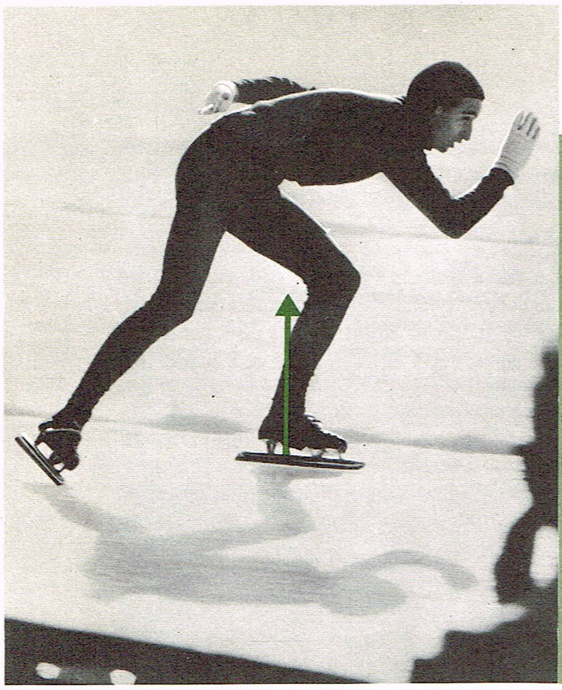
les systèmes matériels

Jusqu'à présent, les mobiles considérés ont été repérés comme des points, à l'aide par exemple de trois coordonnées cartésiennes. Cela signifie soit que nous avons négligé les dimensions du mobile, soit que nous nous sommes bornés à en étudier le mouvement d'un point central, en négligeant le mouvement de rotation qu'il pouvait avoir autour de ce point. Cette conception est évidemment très insuffisante, et il nous faut étendre les lois de la mécanique à tout corps ou même à tout système de corps.

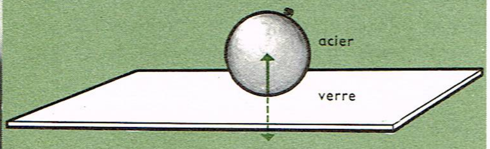
Nous envisagerons principalement des corps solides : qu'ils soient en mouvement ou au repos et quelles que soient les forces qui leur sont appliquées, les distances de deux de leurs points demeurent pratiquement invariables. Nous savons, en effet, que tous les corps se déforment sous l'action des forces appliquées, mais il y a des corps pour lesquels les déformations sont très faibles. On peut admettre que, pour eux, la déformation sous l'action d'une certaine force est localisée au voisinage du point d'application de cette force et négliger cette déformation dans l'étude de leurs mouvements.

actions de contact

Lorsque deux solides sont au contact et qu'ils s'appuient l'un sur l'autre, ils se déforment localement, et des actions de répulsion, empêchant la pénétration d'un corps dans l'autre, se développent entre les particules au contact, sur une petite surface. Ces actions de contact constituent une catégorie très importante de forces, en fait la presque totalité. En dehors d'elles, il n'y a que les actions à distance, dont la nature ne peut être que gravitationnelle ou électromagnétique (et, si l'on fait de la mécanique dans un référentiel non absolu, les forces fictives).

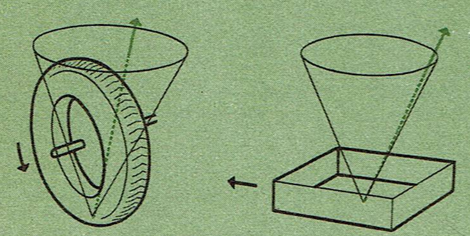


Un patineur.



Bille sur une plaque de verre.

Si les corps sont polis, qu'il y ait mouvement ou non, la réaction est normale aux surfaces en contact.

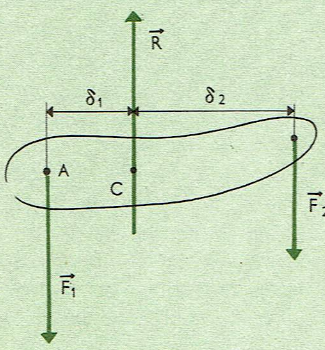


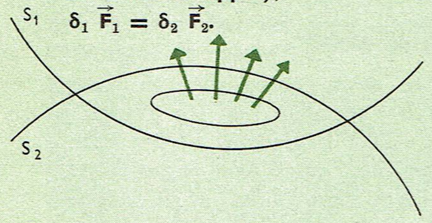
Si les corps sont rugueux, s'il n'y a pas glissement, la réaction est à l'intérieur d'un cône appelé « cône de frottement » (cas des roues d'une voiture). S'il y a glissement, la réaction est sur le cône (cas d'une caisse que l'on traîne).

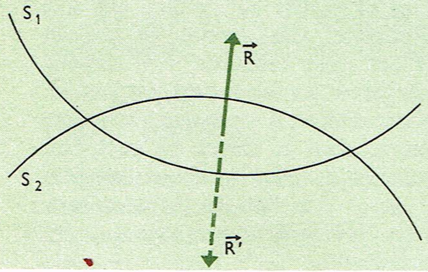


Equilibre du levier.

Écrivons les conditions de l'équilibre en prenant des moments par rapport à C, de manière à éviter d'avoir à considérer le moment de la réaction
$\vec{F}_1 + \vec{F}_2 + \vec{R} = O$ (ce qui détermine la réaction dont on n'a besoin que si l'on manque de confiance dans la solidité de l'appui);
$\delta_1 \vec{F}_1 = \delta_2 \vec{F}_2$.







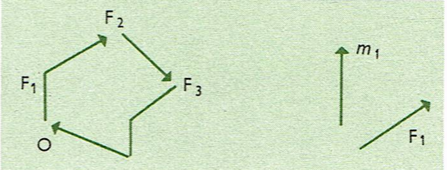
S_2 exerce sur S_1 un ensemble de forces. Si l'on néglige la surface de la zone de contact, l'action de S_2 sur S_1 est la réaction $\vec{R}$, issue du point de contact et dirigée vers l'intérieur de S_1. Inversement (principe de l'égalité de l'action et de la réaction), S_1 exerce sur S_2 une réaction $\vec{R}'$ directement opposée.

solides en équilibre

Un corps solide, soumis à un certain nombre de forces $\vec{F}_1$, $\vec{F}_2$, ..., appliquées en des points P_1, P_2, ..., peut être en équilibre (par rapport à un certain référentiel).

L'étude expérimentale de l'équilibre conduit aux résultats suivants :

La somme géométrique du système de forces $\vec{F}_1$,... est nulle; cela signifie que si, à partir d'un point quelconque O, on met bout à bout des vecteurs équipollents à $\vec{F}_1$, $\vec{F}_2$,... on retombe au point O lorsqu'on a épuisé toutes les forces;



Le moment résultant du système de forces $\vec{F}_1$, ... par rapport à un point déterminé O est nul; cela signifie que si l'on prend les moments par rapport au point O de toutes les forces du système, ils auront une résultante nulle. O est un point fixe quelconque; on démontre que si la somme géométrique des forces est nulle, la construction de leur moment résultant en n'importe quel point donne le même résultat.

dynamique du solide

Étudions maintenant le mouvement (par rapport à un certain référentiel) d'un solide. Pour chaque particule P de masse m constituant le solide, nous pouvons écrire l'identité fondamentale

$$(1) \qquad \vec{F}_P = \frac{d}{dt}\left[\sum m\,\vec{V}_P\right]$$

$\vec{F}_P$ est la force totale appliquée à P. Elle comprend les actions de contact, exercées par d'autres corps appuyant en P sur le solide, les actions à distance et, éventuellement, les forces fictives. Ce sont là les **forces extérieures.** Il faut encore comprendre dans $\vec{F}_P$ les forces exercées sur P par les autres particules du solide, actions de contact qui constituent les forces de cohésion du solide et lui assurent son indéformabilité, actions à distance entre deux particules du solide, généralement négligeables.

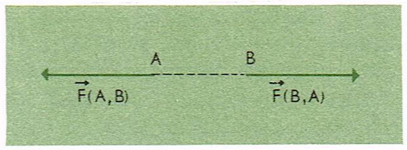
On sait bien peu de choses sur ces forces dites **intérieures,** en dehors du fait qu'elles sont deux à deux opposées (égalité de l'action et de la réaction). Heureusement, cela suffit à établir de nombreux

Phot. Office du tourisme suisse et A. F. P.

résultats, indépendants de l'intensité des forces, ce que nous allons présenter maintenant.

quantité de mouvement d'un solide

Si l'on effectue la somme géométrique de toutes les égalités, il ne figurera au premier membre que la somme géométrique des seules forces extérieures. En effet, les forces intérieures, deux à deux directement opposées, ont une somme géométrique nulle.



Les forces (intérieures) qu'exercent A sur B et B sur A ont une résultante générale et un moment résultant nuls :

$\vec{F}(A, B) + \vec{F}(B, A) = 0,$

$\overrightarrow{OA} \wedge \vec{F}(A, B) + \overrightarrow{OB} \wedge \vec{F}(B, A) = 0.$

Il est naturel d'appeler « quantité de mouvement » d'un solide la somme des quantités de mouvement de chacune des masses qui le constituent; alors,

$$\sum \vec{F}_e = \frac{d}{dt}\left[\sum m \vec{V}_P\right].$$

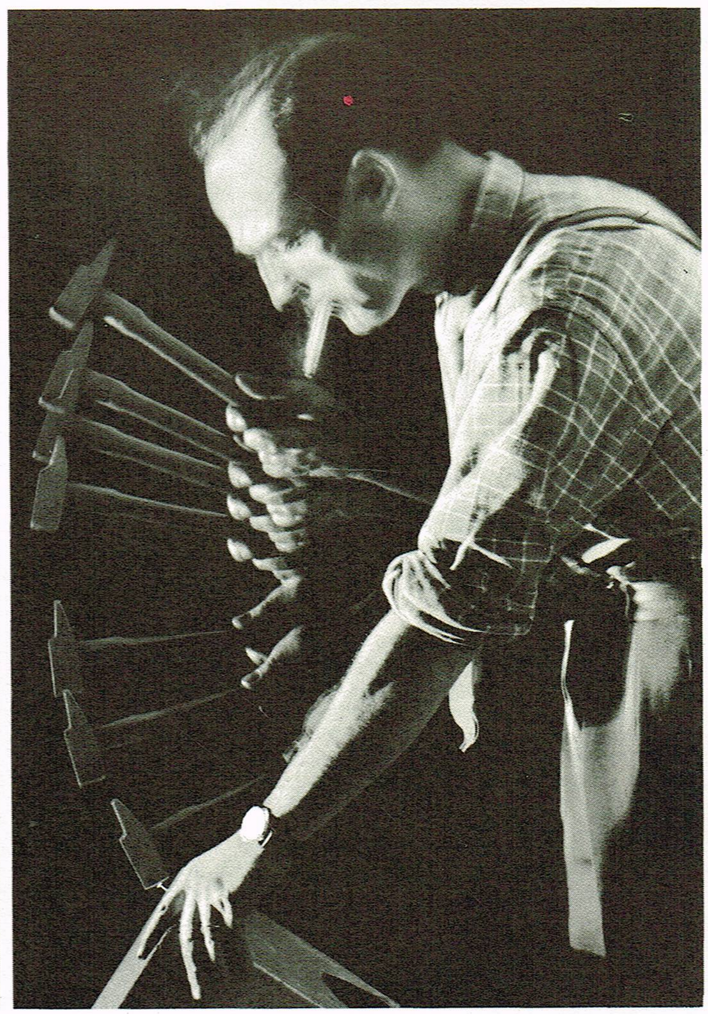
La dérivée de la quantité de mouvement est à chaque instant égale à la somme des forces extérieures.

Ce résultat, établi pour le mouvement d'un solide, s'étend au mouvement d'un

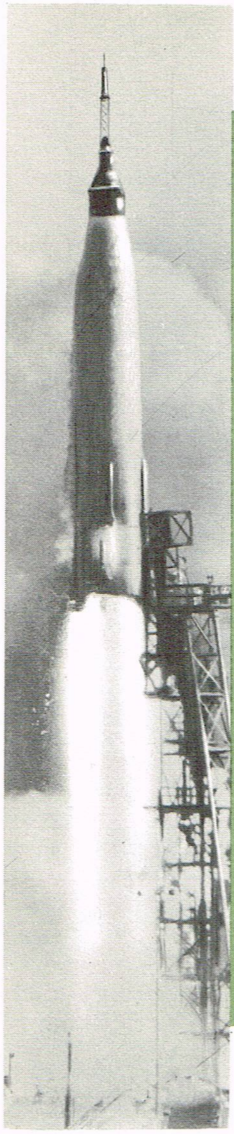
MÉCANIQUE EMPIRIQUE

Masse, poids, force, accélération, inertie... Que de têtes de chapitre de la mécanique réunies dans ce geste banal du menuisier !
Il s'agit d'enfoncer le plus grand nombre de clous avec un minimum de temps et un maximum... d'assurance. D'où l'utilité de la strobographie, qui décompose le geste par une succession de vues prises sur un même cliché à des intervalles de temps réguliers.
L'une des deux photographies concerne un menuisier professionnel; l'autre, un bricoleur.
Cigarette aux lèvres, détendu, sûr de lui-même, l'homme de droite imprime au marteau un mouvement régulier.
A gauche, en revanche, le marteau semble hésiter... C'est pourtant là que se trouve le professionnel. En fait, la main du menuisier prend d'abord la mesure (jusqu'à la septième vue), puis, sans la moindre hésitation, elle fait foncer le marteau sur le clou, tellement vite qu'aucune autre vue intermédiaire n'a pu être prise (l'avant-dernière vue montre le rebondissement du marteau).
Quant au coup de marteau du document de droite, la faible ampleur du mouvement, sa régularité apparente, ne font que traduire le manque d'assurance du menuisier amateur, bien que ce dernier ait mis la pièce dans la meilleure position et pris appui avec sa main gauche.

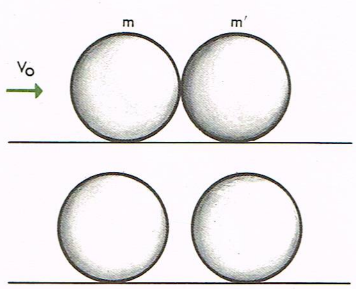




Strobophotographie de J. Salesse-Lavergne.



La poussée est égale au produit de la vitesse d'éjection par la masse éjectée par seconde : 200 kilogrammes de gaz éjectés par seconde à la vitesse de 3 000 m/s. communiquent à la fusée une poussée de 6.10^5 newtons (env. 60 tonnes-force). En vol apparaîtra la nécessité d'éjecter le gaz brûlé à *très grande vitesse.*



Choc de deux billes. Les impulsions proviennent d'une action de contact, donc la quantité de mouvement de l'ensemble des deux billes est conservée par le choc :

$$mv_0 + 0 = mv_1 + m'v'_1.$$

Si les chocs sont élastiques, il faut joindre l'équation $v'_1 - v_1 = v_0 - 0$, d'ou $v_1 = \frac{m - m'}{m + m'} v_0$, qui est nul si $m = m'$ (le mouvement de la première bille se communique à la seconde).

Si le choc est mou, $v_1 = v'_1 = \frac{mv_0}{m + m'}$

système matériel général constitué par plusieurs corps solides, se touchant ou non, et même par des corps déformables. Les forces intérieures, qui sont alors éliminées, comprennent dans ce cas les actions, les uns sur les autres, de tous les corps constituant le système.

comment se propulsent les fusées ?

S'il n'y a pas de forces extérieures, ou que l'on puisse les négliger par rapport aux forces intérieures considérables qui se produisent dans certains cas, la quantité de mouvement totale du système demeurera constante. C'est le cas de la fusée, dont nous négligerons le poids en regard des forces intérieures considérables développées par les gaz de combustion. A un instant donné, elle a une vitesse v (par rapport à la Terre) et une masse M (combustible compris), et elle éjecte par seconde une masse m de gaz avec la vitesse v' (par rapport à la fusée). Une seconde plus tard, elle a une vitesse $v + dv$, mais une masse $M - m$, et la masse m de gaz a la vitesse $v - v'$. La quantité de mouvement, qui était Mv, est devenue

$$(M - m)(v + dv) + m(v - v'),$$

soit approximativement

$$M(v + dv) - m v'.$$

La quantité de mouvement totale demeurant constante, on a donc $Mdv = m v'$. La variation par seconde de la quantité de mouvement de la fusée, c'est-à-dire la force propulsive, ou encore la poussée, est égale au produit de la masse de gaz éjectée par seconde par la vitesse d'éjection.

moment cinétique d'un solide

Écrivons, pour chaque particule P de masse m constituant un solide en mouvement, le résultat relatif à son moment cinétique $\overrightarrow{O\sigma_P}$ par rapport à un point O obtenu précédemment :

$$(2) \qquad \overrightarrow{OP} \wedge \vec{F}_P = \frac{d}{dt}(\overrightarrow{O\sigma_P}).$$

Ici encore l'addition géométrique de toutes les égalités (2) va éliminer les forces intérieures, qui, deux à deux directement opposées, ont un moment résultant nul par rapport à tout point :

$$\sum \overrightarrow{OP} \wedge \vec{F}_e = \frac{d}{dt}\left[\sum \overrightarrow{O\sigma_P}\right].$$

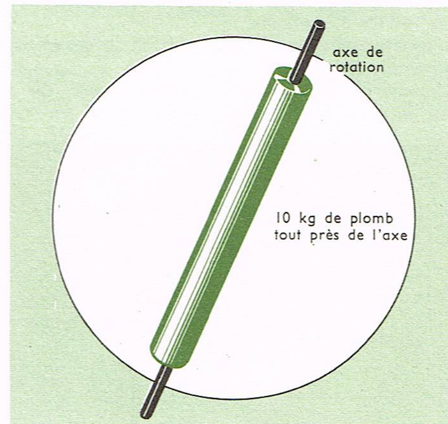
Si l'on appelle **moment cinétique du solide** par rapport au point O la résultante des moments cinétiques de ses masses élémentaires, on peut alors énoncer le résultat suivant :

La dérivée du moment cinétique par rapport à un point fixe O est égale au moment résultant par rapport à ce point des forces extérieures.

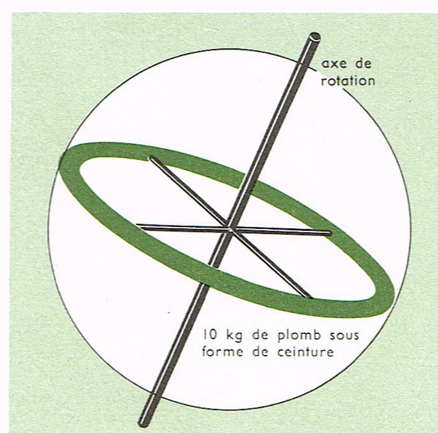
Ce résultat, établi pour le mouvement d'un solide, s'étend au cas du mouvement d'un système matériel général.

comment mesurer l'inertie dans une rotation ?

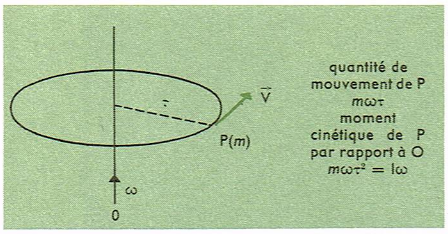
De deux masses de plomb de formes différentes représentées ci-dessous, de 10 kg chacune, la plus difficile à mettre en état de rotation et, plus généralement, d'accélérer angulairement autour de son axe, sera la seconde; ce sera aussi elle qu'il sera le plus difficile d'arrêter; la distance à l'axe intervient dans l'inertie qui oppose les solides à la rotation. Cette propriété est largement utilisée dans les moteurs, où le volant régularise les variations trop brutales dans les rotations.



On fera tourner beaucoup plus facilement l'appareil ci-dessus malgré l'égalité des masses : leur distance à l'axe intervient dans leur inertie à la rotation.

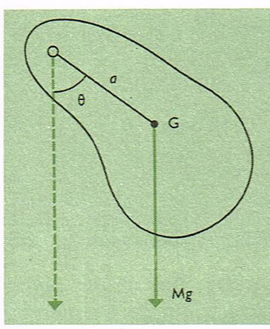


Plus exactement, le moment, par rapport à un point O de l'axe, de la quantité de mouvement d'un point P est porté par l'axe et a pour grandeur $m \omega r^2$.



Lorsqu'on fait la somme, on obtient un vecteur (moment cinétique du corps solide) porté par l'axe et de mesure I ω, $I = \Sigma mr^2$ étant le **moment d'inertie** du solide par rapport à l'axe.

Dans ce cas, la loi du mouvement sera régie par la proposition suivante, qui montre bien comment l'inertie du solide à la rotation est caractérisée par son moment d'inertie I :
Le produit du moment d'inertie I par l'accélération angulaire $\frac{d\omega}{dt}$ est égale au moment des forces extérieures par rapport à l'axe de rotation, c'est-à-dire à la projection sur l'axe du moment des forces extérieures par rapport au point O de l'axe.



Pendule composé.

Mouvement d'un corps pesant mobile sans frottement autour d'un axe horizontal fixe.

Le théorème du moment cinétique utilisé dans les conditions indiquées élimine les actions de contact de l'axe sur le pendule, et les seules forces extérieures à considérer sont les forces de pesanteur.

$$I \frac{d\omega}{dt} = -Mg\,a \sin\theta$$

ou $\frac{d^2\theta}{dt^2} + \frac{Mga}{I} \sin\theta = 0$ **: le pendule composé oscille comme un pendule simple de longueur** $\frac{I}{Ma}$.

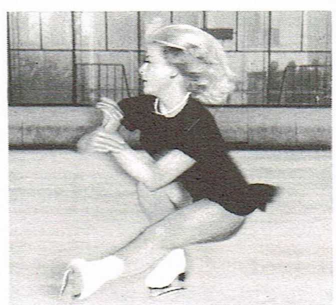
un solide possède aussi de l'énergie

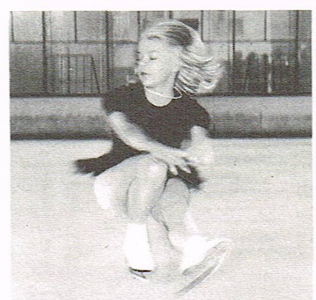
Tout comme dans le cas où les dimensions du corps étaient négligées, un corps solide, et, plus généralement, un système matériel quelconque, possède de l'énergie.

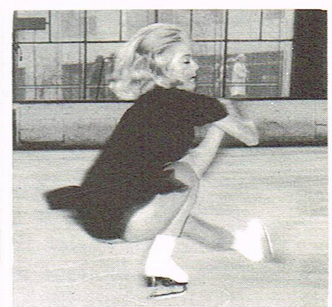
Phot. Larousse.

La variation de la vitesse angulaire de la patineuse s'explique par simple application du théorème du moment cinétique, valable même pour un corps susceptible de se déformer sous l'action de forces intérieures (toujours opposées deux à deux).











Aucune force extérieure n'est à considérer, le poids étant équilibré par la réaction du sol. Le moment cinétique de la patineuse (*) reste donc constant. Mais elle peut, en écartant les bras ou en s'accroupissant et en tendant une jambe, augmenter son moment d'inertie, ce qui a pour effet de diminuer la vitesse angulaire, puisque le produit de ces deux quantités reste constant. Si, au contraire, la patineuse se remet debout et les bras le long du corps, ses masses sont au plus près de l'axe et son moment d'inertie le plus petit possible; alors, la vitesse angulaire prend sa plus grande valeur.

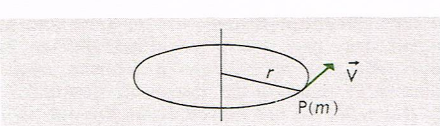
(*) Sur nos photos, Mlle Geneviève Burdel, championne de France junior 1963.

Le procédé général de définition et d'évaluation de l'énergie totale d'un système matériel va être la sommation, si l'on admet cette règle naturelle que l'énergie du tout est la somme des énergies des parties. Naturellement, on vérifiera, dans la limite de la précision des mesures faites, l'accord entre l'énergie ainsi calculée et l'énergie que l'on pourra mesurer.

Suivant le mouvement du solide, l'évaluation de l'énergie sera plus ou moins simple. Bornons-nous donc à quelques exemples.

Un train se déplace sur une portion rectiligne de rail : toutes les vitesses de toutes les parties du train sont, à tout instant, équipollentes (nous négligeons ici les pièces mobiles par rapport au train, roues, rotor du moteur de la locomotive, etc.). Chaque masse m a la vitesse V du train, donc une énergie cinétique $\frac{1}{2} mV^2$, et l'ensemble du train a l'énergie cinétique $\frac{1}{2} MV^2$, M étant la masse totale du train. L'énergie cinétique d'un solide en translation est égale au demi-produit de sa masse totale par le carré de la vitesse de translation. C'est ainsi qu'on évaluera l'énergie cinétique d'un corps tombant en translation dans le champ de pesanteur terrestre ou l'énergie cinétique d'un satellite artificiel dont les hublots ou les antennes pointent constamment vers les mêmes étoiles fixes. Naturellement, l'énergie cinétique du train et celle du corps en chute libre seront calculées par rapport au référentiel Terre, celle du satellite artificiel par rapport au référentiel Terre-Étoiles.

Prenons un autre cas simple : celui d'un solide en mouvement de rotation autour d'un axe fixe. Nous connaissons la distribution des vitesses des différents points de ce solide. Ici, le fractionnement est nécessaire, ainsi que le calcul de l'énergie cinétique de chaque partie avant sommation. Pour un élément de masse situé à la distance r de l'axe, la



vitesse vaut ωr; l'énergie cinétique est donc $\frac{1}{2} m\omega^2 r^2$. Mais la vitesse de rotation ω est la même pour chaque partie du solide, du fait de sa rigidité. C'est



Il y a potentiel pour les forces de pesanteur : l'eau retenue par le barrage possède une énergie potentielle susceptible de se transformer en énergie cinétique ou de produire du travail.

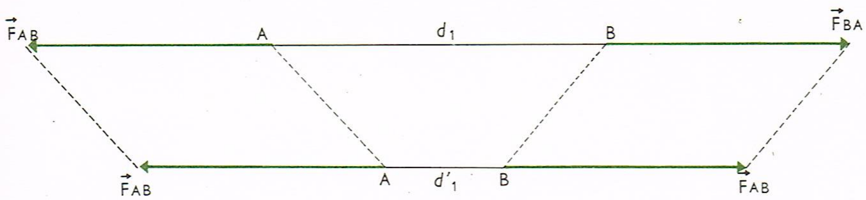
donc les produits de la masse par le carré de la distance à l'axe qu'il nous faut sommer. Nous retrouvons le moment d'inertie I du solide par rapport à son axe de rotation. Pour un solide en rotation autour d'un axe fixe, l'énergie cinétique est le **demi-produit de son moment d'inertie par rapport à son axe par le carré de la vitesse angulaire de rotation.**

énergie cinétique et travail, énergie potentielle

Pour un corps ponctuel, nous avons vu que la variation de l'énergie cinétique était égale au travail de la force totale appliquée. Par sommation, nous aurons un résultat analogue pour un système matériel quelconque : la variation de l'énergie cinétique du système est égale à la somme des travaux de toutes les forces qui lui sont appliquées. Cet énoncé, qui constitue le théorème de l'énergie, fournit une seule équation scalaire, au contraire des deux autres théorèmes généraux, ceux de la quantité de mouvement et du moment cinétique, qui, eux, fournissent chacun une équation vectorielle, c'est-à-dire trois équations scalaires. Mais il faut faire attention : le théorème de l'énergie fait intervenir le travail de toutes les forces appliquées au système, aussi bien les forces extérieures que les forces intérieures, et le travail de ces dernières n'est, en général, ni nul ni même commode à déterminer.

Un cas particulier intéressant est celui où la totalité des forces appliquées au système ne dépend que de la position du système et où, de plus, le travail de toutes ces forces ne dépend que de la configuration initiale et de la configuration finale du système, et non des états intermédiaires. Dans ce cas, l'ensemble de ces forces dépend d'un potentiel. Le système possède une énergie potentielle, et le théorème de l'énergie prend un aspect de théorème de conservation :

Phot. G. Métrailler, Borlat.



Bien que formant un ensemble de vecteurs de résultante et de moment résultant nuls, les forces ont effectué le travail non nul (ici positif) $\mathcal{T} = \vec{F}\,(d_1 - d'_1)$, ce qui rend erronée toute tentative d'appliquer sans précautions tout théorème de conservation de l'énergie : il faut travailler pour dilater un ressort.

l'énergie totale, somme de l'énergie cinétique et de l'énergie potentielle, garde une valeur constante au cours du temps. Cela constitue une importante intégrale première pour le mouvement du système.

L'utilisation des sept équations scalaires (ou deux vectorielles et une scalaire) fournies par les théorèmes de la quantité de mouvement, du moment cinétique et de l'énergie est délicate. Parmi les forces

$$\sum \vec{F} = \frac{d}{dt}\left(\sum m\,\vec{V}_P\right)$$

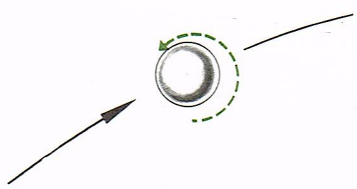
$$\sum \vec{OP} \wedge \vec{F} = \frac{d}{dt}\sum \vec{O}\sigma_P$$

$$\sum \mathcal{T}(\vec{F}) = \Delta E$$

appliquées à un système, il y en a qui sont connues, les forces de pesanteur, par exemple; il y en a qui sont inconnues, les forces de cohésion des solides, par exemple; il y en a qui ne sont que partiellement connues, les actions de contact entre solides, par exemple. Le mouvement du système est déterminé par les forces connues et par les conditions initiales. L'élimination de certaines forces inconnues peut s'effectuer avant la mise en équations en utilisant judicieusement les théorèmes généraux; sinon, il faut procéder à une élimination algébrique souvent pénible après mise en équations. Enfin, l'intégration des équations du mouvement sera facilitée par la connaissance d'intégrales premières, qu'il convient donc de rechercher avec soin.

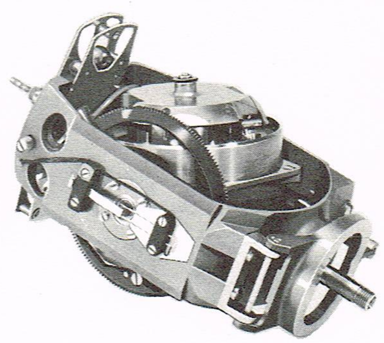
Nous nous contenterons de ce bref aperçu de la mécanique des théorèmes généraux et signalerons simplement son prolongement, la mécanique analytique, élaborée par Lagrange, dont l'importance scientifique dans les théories physiques modernes est considérable. Notre dessein a été, en effet, de montrer comment s'introduisent et entrent en jeu les notions et les principes fondamentaux de la science mécanique, et non pas d'expliciter les procédés techniques de leur utilisation.

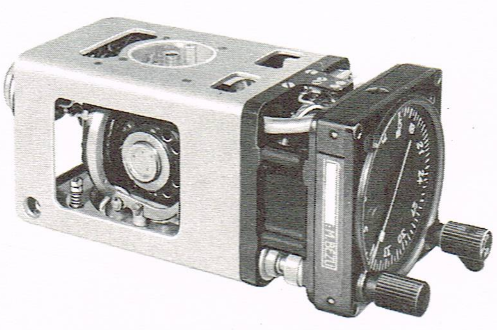
Conséquences du théorème du moment cinétique.



La balle, lancée avec une certaine rotation initiale, conserve sa rotation autour d'un axe qui garde une direction fixe, le moment des forces extérieures par rapport au centre de gravité de la balle étant nul.

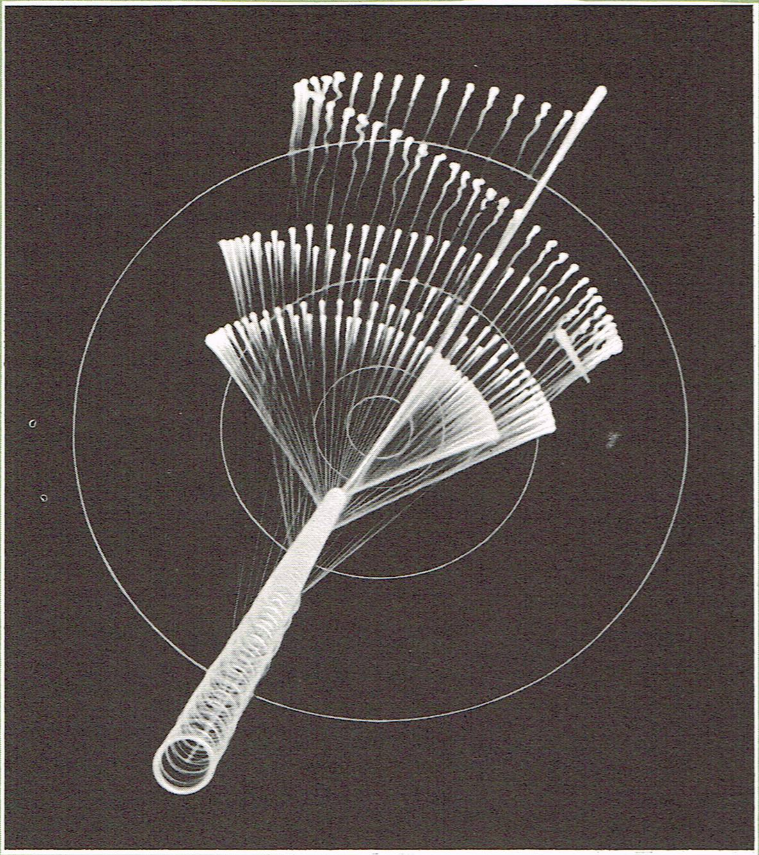
La suspension « à la Cardan » permet de réaliser cet effet en un point fixe. Le gyroscope est mobile autour de son centre des masses maintenu fixe. Lancé avec un seul mouvement de rotation autour de son axe, sa vitesse de rotation (si on néglige les frottements) et la direction de son axe restent invariables. Mais pour de grandes vitesses de rotation, il faudrait tenir compte, sur la Terre, des forces d'inertie de Coriolis. Mieux vaut faire de la Mécanique dans le référentiel « Terre-Étoiles » : la direction de l'axe reste invariable par rapport aux étoiles. D'où la solution idéale du « problème du cap », le gyroscope de Foucault : le gyroscope est lancé avec une grande vitesse de rotation autour de son axe dirigé vers l'étoile polaire. Cet axe indiquera toujours la direction de la Polaire.





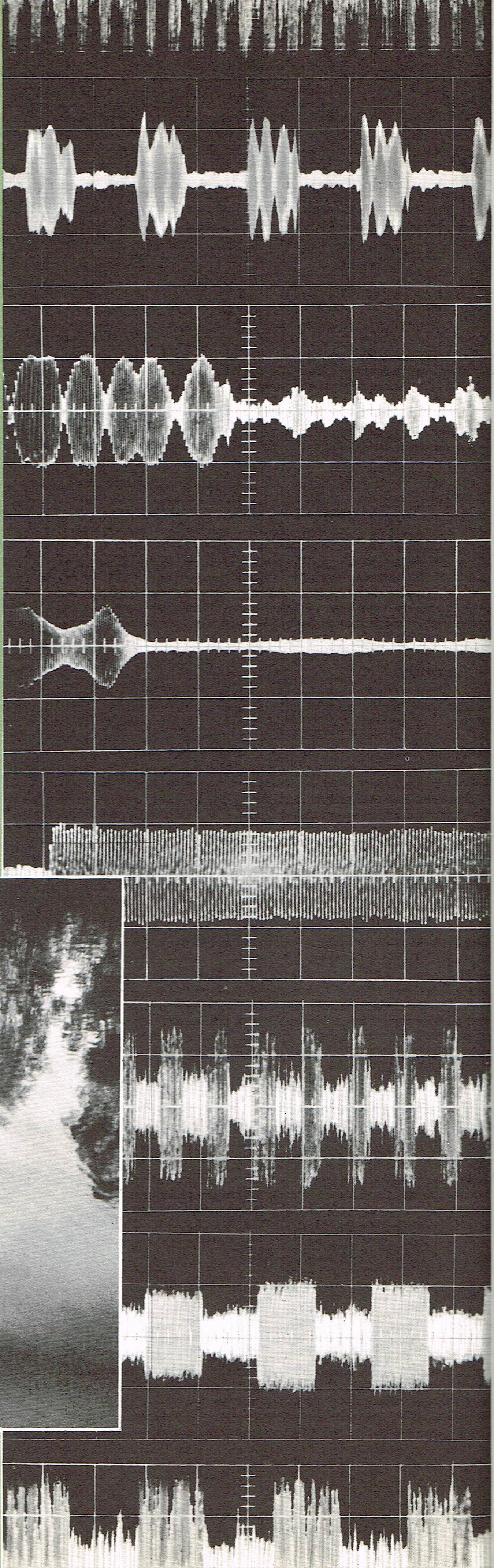
A cause des problèmes que posent la suspension et ses frottements inévitables, l'entretien de la rotation et le recalage nécessaire de l'instrument, cette solution idéale est remplacée par l'utilisation d'appareils complexes : le compas gyromagnétique et le compas gyroscopique, qui ne sont utilisables qu'en des lieux éloignés du pôle, et enfin par le « directionnel », qui peut servir aux avions survolant le pôle, ou aux sous-marins qui atteignent le pôle au-dessous de la banquise.

Phot. S. F. E. N. A. et Bezu.

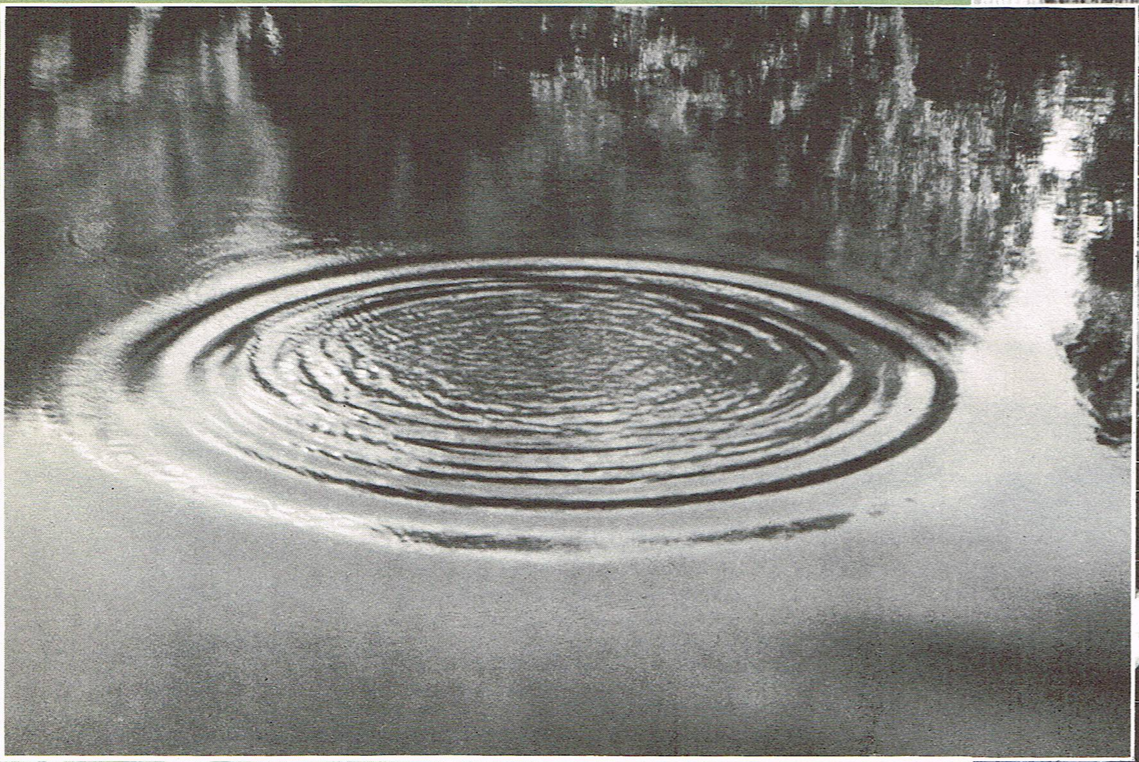


Ballet du spot d'un radar volontairement excentré, affolé, pour les besoins de la photographie.

Le monde des ondes dans sa diversité.



Langage étonnamment varié et complexe que celui du monotone « bip-bip » du « Spoutnik », traduit ici en oscillogrammes que, seuls, les possesseurs du code pourront interpréter.



Cercles providentiels pour la compréhension des premières notions de la physique des ondes à l'école.

Phot. C. S. F., Doc. Philips, André Laporte.

la physique des ondes

par Pierre MESNAGE et Jean-Charles VIÉNOT

Le monde de la mécanique, où les seuls attributs des corps sont, avec la position dans l'espace, l'énergie et la quantité de mouvement, séduit par son ordonnance rigoureuse. C'est l'univers rêvé par Descartes; pour la plupart des penseurs du « Siècle des lumières », tels Lagrange ou d'Alembert, il englobe toute la réalité. En même temps, toutefois, il rebute par son aridité. Ce n'est qu'un squelette d'univers, une scène vide où les acteurs de la « comédie aux cent actes divers » des phénomènes physiques n'ont pas encore pris place. De ces acteurs, les premiers sont les phénomènes électriques.

Nous avons conscience de mettre en jeu des actions électriques à chaque instant de notre vie quotidienne, mais peu d'entre nous sont attentifs au rôle essentiel de l'électricité dans les structures les plus intimes et les plus secrètes de la matière. C'est pourquoi le lecteur est ici convié à une exploration très rapide des lois les plus importantes des phénomènes électriques, électromagnétiques et optiques; exploration ardue dans un terrain mouvementé et complexe. Il ne regrettera cependant pas les efforts qui vont lui être demandés, car les perspectives aperçues — ou même seulement entrevues — en valent la peine.

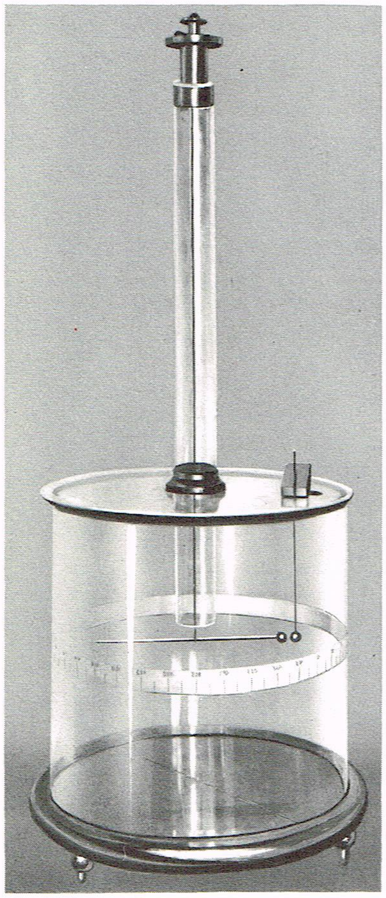
où l'on voit apparaître l'électrisation

La première étape sera la présentation des actions électriques sous leur forme la plus simple et pour ainsi dire la plus pure : les forces électrostatiques.

Pour cela, nous remonterons très brièvement aux origines : l'attraction des corps légers par l'ambre préalablement frotté, qui aurait été découverte par Thalès de Milet un certain nombre de siècles avant notre ère. Le fait de frotter l'ambre lui confère un état particulier le rendant capable d'attirer des brins de plume ou des parcelles de liège; cette force a ce caractère singulier qu'elle s'exerce à distance, sans contact. Nous faisons involontairement l'expérience quand le soir jaillissent des étincelles de nos vêtements en Rhovyl. Le premier pas dans la connaissance d'un phénomène étant de le nommer, celui-là a tout naturellement reçu le nom d'« électrisation» (*elektron* signifie « ambre », en grec). Cette propriété de l'ambre est restée bien longtemps isolée et inexploitée. L'observation s'est, cependant, peu à peu précisée, et l'on s'est aperçu qu'il y avait des forces de deux sortes : d'attraction et de répulsion. C'est-à-dire qu'il y avait deux types et deux seulement d'électrisation : après avoir subi l'électrisation de même type, deux objets se repoussent; électrisés chacun selon un type différent ou opposé, ils s'attirent.

Franchissons les siècles à tire-d'aile, et présentons-nous devant Coulomb (1788) pour apprendre de lui comment cette force varie avec la distance.

Sa balance de torsion lui permit de reconnaître que la répulsion entre deux corps électrisés de la même manière varie comme l'inverse du carré de la distance. L'attraction entre les corps électrisés de manière opposée suit la même loi. Dès lors, la **loi de Coulomb** et son application aux deux sortes de charges électriques, que nous appelons aujourd'hui « positive » et « négative », sont des acquisitions définitives de la science. Ainsi, deux charges q et q' placées à distance r l'une de l'autre exercent mutuellement l'une sur l'autre une force f donnée par l'expression $f = K\frac{qq'}{r^2}$, K étant une constante dépendant des unités choisies. Dans le système M. K. S. A. (système cohérent dont les unités fondamentales sont le mètre, le kilogramme, la seconde et l'ampère), K vaut 9.10^9. Deux gouttes sphériques d'un millimètre de rayon, chargées aussi fortement qu'on peut le faire dans l'air atmosphérique (la déperdition des charges en raison d'un isolement imparfait impose, en effet,



Balance de Coulomb.
Coulomb vérifie que des corps électrisés de la même manière se repoussent selon une force $f = K\frac{qq'}{r^2}$.
Des corps électrisés de manière opposée s'attirent selon une loi analogue (1788).

une limite pratique à l'électrisation) et placées à un centimètre l'une de l'autre, exercent l'une sur l'autre une force égale à une dyne (milligramme-force). Cela paraît bien peu, mais si l'on pouvait amener deux charges de 1 coulomb (1) à 1 centimètre l'une de l'autre (ce qui n'est guère réalisable, sauf peut-être fugitivement, au sein des explosions atomiques), l'attraction ou la répulsion atteindrait alors 9 milliards de tonnes-force. La physique exige une intense gymnastique intellectuelle pour se familiariser avec les chiffres (2).

Signalons également qu'en employant la même méthode et le même instrument Coulomb découvre que la même loi régit aussi bien la répulsion ou l'attraction de deux pôles d'aiguilles aimantées. Similitude à la fois suggestive et trompeuse, qui a fourni un cadre commun au développement des théories de l'électricité et du magnétisme, et en même temps a fait peser une lourde hypothèque sur l'intelligence profonde des phénomènes, comme nous aurons lieu de le constater.

Pour essayer de nous faire une image aussi juste et vivante que possible des forces et des charges électriques, il ne convient pas de faire ici un exposé déductif de l'électrostatique à partir de la loi de Coulomb (ce qui serait parfaitement logique, mais aussi pesant qu'ennuyeux), mais plutôt de décrire avec quelques détails les expériences de Millikan, d'où nous tirerons les notions également fondamentales de champ électrique et d'électron (quantum d'électricité).

l'expérience de Millikan

Naturellement, on n'avait pas attendu Millikan (1910) pour parler d'électrons, ni les années voisines de 1900 pour parler de champ électrique : l'idée première en remonte à Faraday (1840), et l'étude mathématique approfondie à Maxwell (1873). C'est seulement au terme de cet exposé, lorsqu'il sera question des « équations de Maxwell », que nous pourrons mesurer l'importance de la notion de champ (déjà utilisée en mécanique) et voir combien son application a rénové une physique bâtie jusqu'alors sur l'idée de force. En revanche, les expériences que nous allons voir ont puissamment contribué à imposer l'idée que la charge électrique est toujours constituée de grains indivisibles. Le nom d'« électrons » avait été forgé pour eux avant même que leur existence eût été universellement reconnue.

Pour expérimenter sur les forces électriques, il faut d'abord des corps électrisés : dans les expériences de Millikan, ce sont des gouttelettes d'huile produites par un pulvérisateur A, assez semblable à un pistolet à peindre.

Pourquoi des gouttes d'huile? Parce qu'elles ne s'évaporent pas et gardent ainsi une masse constante. En outre, elles s'électrisent fort bien. Comment? Par frottement à la sortie du pulvérisateur.

Abandonnées à elles-mêmes dans l'air, ces petites sphères tombent comme n'importe quel objet, et rien ne montre alors qu'elles portent une charge électrique. En raison de la résistance de l'air, leur vitesse de chute devient presque immédiatement uniforme, et cette vitesse est proportionnelle à leur poids, nous voulons dire à la force exercée par le champ de pesanteur.

Les choses changent quand les gouttes électrisées sont mises en présence d'autres charges électriques. Mais, dira-t-on, n'exercent-elles pas déjà des actions les unes sur les autres ? Certes oui, en principe, mais si le brouillard est assez dilué, les gouttes sont éloignées les unes des autres, et leurs actions réciproques négligeables. D'un tout autre ordre de grandeur est l'action des charges portées par deux plateaux métalliques parallèles M et N, reliés aux pôles d'une batterie d'accumulateurs formée d'éléments suffisamment nombreux pour fournir une tension élevée. **Dès qu'une goutte pénètre dans l'espace entre les plateaux, sa vitesse de chute est brusquement modifiée, ce qui manifeste l'entrée en jeu d'une force** (électrique), superposée à la pesanteur. On mesure cette vitesse de chute avec le plus de soin et de précision possible, en chronométrant le temps que met la goutte à traverser le champ du viseur d'observation. De la comparaison des vitesses sont tirées toutes les conclusions.

(1) Qu'est-ce au juste que le coulomb? Une réponse tout à fait complète demanderait que l'on explique la formation des unités dans un système cohérent. En un certain sens, le coulomb est une unité fondamentale, donc arbitraire comme l'est la mesure du kilogramme-étalon. Trouvera-t-on plus satisfaisante sa définition électrodynamique? Si oui, c'est la définition de l'ampère (coulomb par seconde) qui s'en déduit : on imagine deux fils parallèles tendus indéfinis, distants de 1 mètre, parcourus par un même courant de 1 ampère; la force d'attraction qu'ils exercent l'un sur l'autre vaut 2.10^{-7} newton par mètre de longueur de fil.

(2) Neuf milliards s'écrit 9.10^9 (« dix puissance neuf »); en effet, il faut écrire neuf zéros après le 9 pour représenter, en système décimal, un tel nombre (9 000 000 000). De même, cent s'écrit 10^2; mille, 10^3; un million, 10^6; cinquante millions, 50.10^6 (ou 5.10^7). L'inverse (0,1) de dix s'écrit 10^{-1}; et, de même, un centième, 10^{-2}, un millième, 10^{-3}, un millionième, 10^{-6}, un milliardième, 10^{-9}. A défaut de représentation, impossible pour l'esprit, l'écriture de très grands ou de très petits nombres est donc possible.

Phot. J. Boyer.

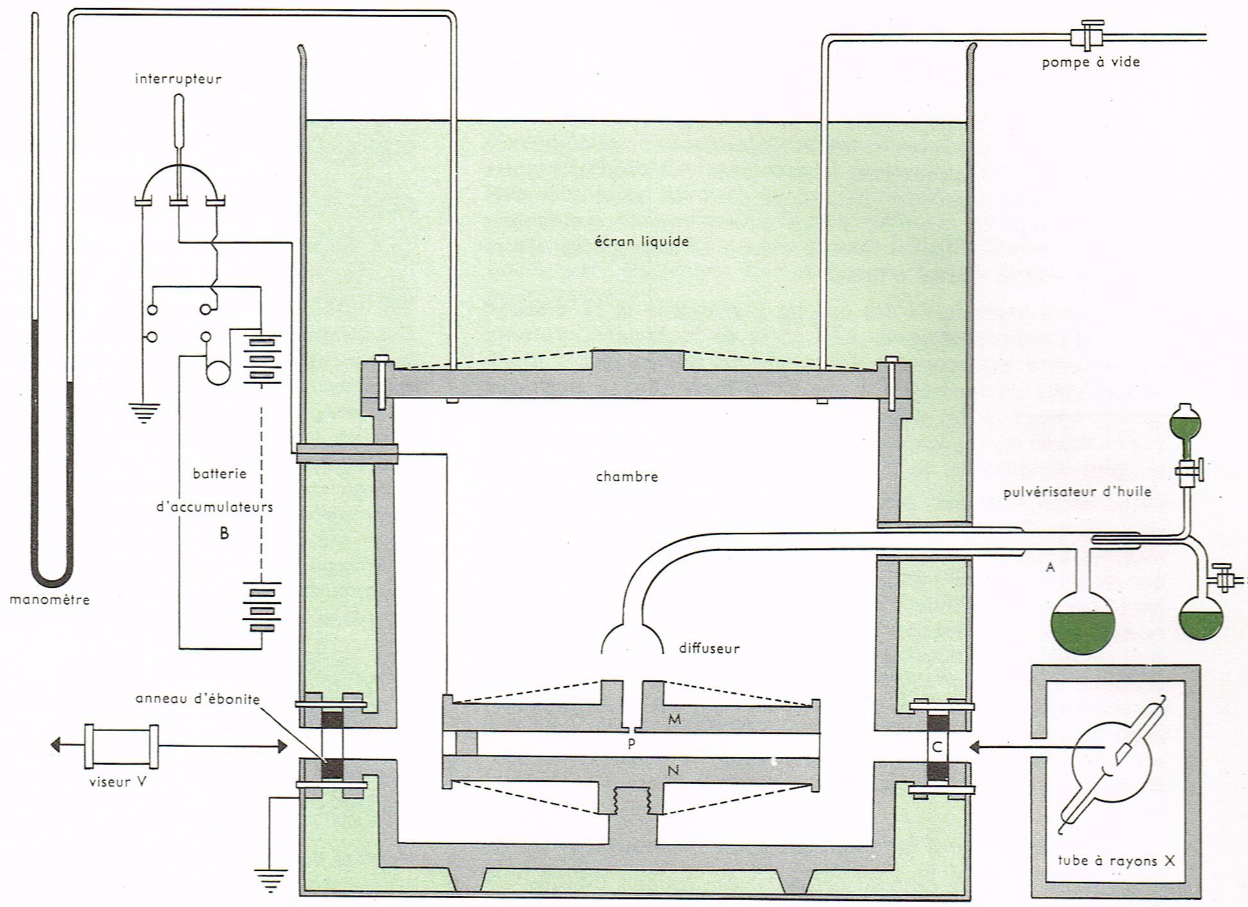


Schéma de l'expérience de Millikan.

On remarque notamment le pulvérisateur d'huile A, les plateaux M et N reliés électriquement à la batterie B, le viseur V à travers lequel on observe l'enceinte *p*. Dès qu'une goutte pénètre dans l'espace entre les deux plateaux, sa vitesse de chute est brusquement modifiée, ce qui manifeste l'entrée en jeu d'une force.

Le livre l'*Électron*, de Millikan, publié à Chicago en 1917, et traduit en français en 1926, est un grand classique de la physique. Sa lecture après quarante ans n'a aucunement perdu de son intérêt.

« L'on purifiait de ses poussières l'air soufflé avec ce brouillard, en le faisant passer dans un tube contenant du coton de verre. Les minuscules gouttelettes d'huile dont était formé le brouillard (la plupart avaient un rayon de l'ordre du millième de millimètre) tombaient lentement dans la chambre C et, de temps en temps, il s'en trouvait une qui passait par le minuscule trou d'aiguille p *situé au milieu du plateau circulaire en laiton M, de 22 centimètres de diamètre, qui formait l'un des deux plateaux du condensateur à air. Trois colonnes d'ébonite maintenaient l'autre plateau N à 16 millimètres au-dessous du premier. L'on chargeait ces plateaux au moyen de l'interrupteur I, l'un positivement et l'autre négativement, en les réunissant aux bornes d'une batterie d'accumulateurs B de 10 000 volts ; en tournant l'interrupteur de l'autre côté vers la gauche, l'on mettait en court-circuit les plateaux, ce qui annulait le* **champ électrique.** *Un faisceau intense de lumière qui passait par des fenêtres diamétralement opposées, percées dans l'anneau d'ébonite* c, *éclairait des gouttelettes d'huile après leur passage à travers* p. *Quand on regardait dans une troisième fenêtre pratiquée dans* c, *du côté de l'observateur, une gouttelette qui venait de passer, celle-ci apparaissait comme une brillante étoile sur un fond obscur. Les gouttelettes qui passaient par* p *étaient en général très fortement chargées par suite du frottement qui se produisait quand on soufflait le brouillard ; c'est pourquoi, si l'on faisait agir le* **champ** *dans la direction convenable, elles se trouvaient repoussées vers M. Immédiatement avant que la gouttelette puisse atteindre M, on mettait les plateaux en court-circuit, et on laissait la goutte tomber sous l'action de la pesanteur jusqu'à ce qu'elle soit tout près de N ; l'on faisait alors agir de nouveau le* **champ** *pour renverser encore le sens de son déplacement. De cette façon, on obtenait un mouvement constant d'allée et venue de la goutte entre les plateaux. La première fois que l'on fit l'expérience, un ion fut capté en moins de quelques minutes, et sa capture fut signalée à l'observateur par le changement qui se produisit dans la vitesse d'ascension due à l'action du champ. »*

t_g	t_f
13,6	12,5
13,8	12,4
13,4	21,8
13,4	34,8
13,6	84,5
13,6	84,5
13,7	34,6
13,5	34,8

« La colonne « t_g » du tableau ci-contre indique les temps successifs qu'il faut à la gouttelette pour tomber entre les deux fils réticules de la lunette d'observation, la distance de ces fils correspondant à une distance réelle de chute de 0,522 2 cm. L'on verra que ces chiffres sont identiques aux erreurs près que comportent les mesures faites avec un compte-secondes. La colonne « t_f » contient les temps successifs qu'il faut à la gouttelette pour monter sous l'influence du **champ électrique** *produit, qui dans ce cas était dû à une différence de potentiel de 5 051 volts établie entre les plateaux M et N. L'on voit qu'après la seconde ascension l'intervalle de temps passe de 12,4 à 21,8, ce qui signifie, puisqu'il s'agit d'une goutte positive, qu'un ion négatif de l'air a été capté. La durée d'ascension suivante, 34,8, indique la capture d'un second ion négatif. Le nombre suivant, 84,5, résulte encore de la capture d'un nouvel ion négatif. Cette charge se maintient pendant deux ascensions, puis l'on revient à une vitesse de 34,6, indiquant la capture d'un ion positif transportant précisément la même charge que l'ion négatif qui avait provoqué une variation inverse de durée, c'est-à-dire de 34,8 à 84,5. »*

Il y a toujours dans l'air des ions, particules errantes chargées électriquement (si l'on craint qu'il n'y en ait pas assez, quelques traces de produits radio-actifs — ou un tube à rayons X — se chargeront d'en produire, selon un mécanisme que nous verrons plus loin). Si les gouttelettes d'huile sont minuscules à notre échelle, elles sont énormes à l'échelle d'un ion isolé, et ce dernier se comporte vis-à-vis d'elles pratiquement comme une charge électrique dépourvue de masse et susceptible d'être capturée par la goutte d'huile.

On observe alors ceci : dès que les plateaux M et N, d'abord en court-circuit, sont reliés aux pôles de la batterie, l'espace qui est entre eux acquiert une propriété particulière : précisément celle de soumettre à une force toute charge électrique qui s'y trouve. C'est exactement ce qu'on exprime en disant que cette région est devenue le « siège d'un champ électrique », ou, plus brièvement, qu'elle est un **champ électrique.**

Cette notion ne doit pas rester seulement qualitative; nous voulons quelque chose de précis. Les résultats chiffrés des mesures expérimentales (il faut prendre la peine de les regarder un à un) montrent pour une même goutte et une même batterie (c'est-à-dire le champ entre M et N restant le même) différentes vitesses correspondant aux forces exercées sur des charges différentes. Ces forces électriques ont cependant quelque chose en commun, et ce quelque chose est précisément la valeur du champ. Avec le même dispositif, on pourrait aussi (en éliminant les ions errants) mesurer les différentes forces exercées sur une même charge lorsqu'on modifie le nombre des éléments de la batterie; ces forces auraient alors en commun la charge.

Une situation tout à fait analogue a été rencontrée en mécanique à propos de la pesanteur. La force de pesanteur, le « poids », est le produit de deux facteurs, le premier (scalaire : la masse *m*) caractérise le corps subissant l'action, et le second (vectoriel : l'accélération $\vec{g}$ de la pesanteur) exprime la cause qui le produit : $\vec{P} = m\vec{g}$. De même, la force subie par une petite charge électrique dans un champ s'écrit $\vec{F} = q\vec{E}$, où q est la grandeur de la charge et $\vec{E}$ le vecteur « champ ».

Malgré la similitude des mots, pas de confusion à craindre entre le champ-région de l'espace où toute charge introduite subit une force F, et le champ (on devrait dire l' « intensité du champ »)-valeur locale du quotient de la force par la charge d'épreuve : $E = \frac{F}{q}$ (relation entre nombres, entre scalaires et non plus entre vecteurs).

Nous voyons pointer ici une notion qui prendra dans la suite tout son relief et reviendra comme un leitmotiv : celle de **valeur locale, de grandeur locale définissable dans un volume aussi petit qu'on veut, attachée à chaque point, pouvant être suivie lorsqu'on passe d'un point à un point voisin, aussi voisin qu'on veut, en un mot adaptée au puissant instrument qu'est le calcul différentiel.**

Le champ est donc une virtualité qui devient force aussitôt qu'on lui en donne l'occasion en introduisant dans l'espace où il règne un corps portant une « charge d'épreuve » (1).

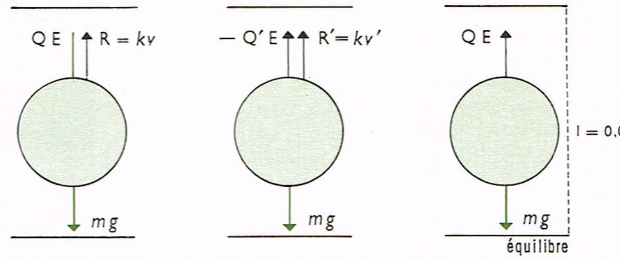
La seule nouveauté, quand on passe du poids à la force électrique, est la possibilité de deux sens opposés pour la force exercée par un même champ, suivant le type d'électrisation du corps. C'est ce qu'expriment les deux signes possibles pour la charge q; si les nombres algébriques n'avaient pas été imaginés depuis longtemps, c'eût été une bonne occasion de les inventer; cependant, il faut rendre cette justice aux mathématiciens que l'emploi spéculatif des créations de leur esprit a précédé l'application pratique de celles-ci aux réalités physiques.

(1) Cela suppose que ni le corps ni les charges d'épreuve ne perturbent le champ de manière appréciable. En théorie, l'introduction de toute charge dans un champ le modifie, de même que la présence d'un caillou modifie le champ de gravitation de la Terre. En fait, il est clair qu'un caillou — ne pesant que quelques grammes — n'apporte qu'une modification infime, au-dessous de toute valeur mesurable. De même, dans l'expérience de Millikan, la perturbation apportée au champ par les quelques gouttes chargées qui pénètrent entre les plateaux est absolument négligeable.

une électricité nécessairement granulaire

Que nous apprennent les expériences de Millikan, telles que les résument les chiffres du tableau ci-dessous sur la charge électrique?

On commence par calculer les vitesses de chute en divisant les longueurs de chute par les temps correspondants (si nous nous sommes crus obligés de les donner avec tant de décimales, ce n'est pas par amour de la précision, mais par exacte soumission à la réalité des faits). Un calcul élémentaire montre que la différence entre les vitesses v de la goutte (lorsque sa charge est Q) et v' de la même goutte (lorsque la charge est devenue Q') est proportionnelle à la différence de charge Q' — Q, c'est-à-dire à la charge qui, captée par la goutte, a fait passer sa vitesse de v à v'.



$v = 0$ (goutte équilibrée) lorsque $mg = qE$, soit $m \sim 3{,}35 . 10^{-15}$ kilogramme
[goutte d'huile de densité 0,8, de rayon 10^{-3} mm]).
$q \sim 1{,}6\ 10^{-19}$ (charge de l'électron en M.K.S.A.)
$g \sim 10\ m/s$ (accélération de la pesanteur)
$E = \frac{mg}{q} = \frac{3{,}35 .\ 10^{-14}}{1{,}6\ 10^{-19}} = 2{,}1 . 10^5$.
La différence de potentiel entre les deux plateaux (champ uniforme) doit être :
$V = El = 2{,}1\ 10^5 \times 0{,}016 = 3\ 350$ volts
si les plateaux sont distants de 16 mm (0,016 m).

Laissons maintenant la parole à l'illustre physicien :

« L'on trouvera dans le tableau ci-dessous les valeurs successives de v et de (v' — v), ces dernières étant obtenues en retranchant l'une de l'autre des valeurs successives de v. »

v et v'	v' — v
$\frac{0{,}522\ 2}{12{,}45} = 0{,}041\ 96$	
	0,041 96 — 0,023 90 = 0,018 06 : 2 = 0,009 03
$\frac{0{,}522\ 2}{21{,}5} = 0{,}023\ 90$	
	0,023 90 — 0,015 05 = 0,008 85 : 1 = 0,008 85
$\frac{0{,}522\ 2}{34{,}7} = 0{,}015\ 05$	
	0,015 05 — 0,006 14 = 0,008 91 : 1 = 0,008 91
$\frac{0{,}522\ 2}{85{,}0} = 0{,}006\ 14$	
	0,015 05 — 0,006 14 = 0,008 91 : 1 = 0,008 91
$\frac{0{,}522\ 2}{34{,}7} = 0{,}015\ 05$	
	0,032 64 — 0,015 05 = 0,017 59 : 2 = 0,008 80
$\frac{0{,}522\ 2}{16{,}0} = 0{,}032\ 64$	
	0,032 64 — 0,015 05 = 0,017 59 : 2 = 0,008 80
$\frac{0{,}522\ 2}{34{,}7} = 0{,}015\ 05$	
	0,023 90 — 0,015 05 = 0,008 85 : 1 = 0,008 85
$\frac{0{,}522\ 2}{21{,}85} = 0{,}023\ 90$	

« La dernière colonne, qui contient les plus grands communs diviseurs des différences $v' - v$, montre qu'aux erreurs près dues au compte-secondes, les charges captées ont toutes la même grandeur, sauf dans trois cas. Dans ces trois exceptions, les charges captées sont exactement le double de celles observées dans les autres cas. Des relations exactement du même genre ont été trouvées sans aucune exception dans toutes les expériences, quelle que fût la nature du gaz dans lequel les gouttes étaient en suspension et quelle que fût l'espèce des gouttes employées pour capturer des ions. On a, maintes fois, pu garder en observation une même goutte pendant cinq à six heures, et l'on a vu capturer non pas huit ou dix ions, comme dans l'expérience ci-dessus, mais des centaines. En somme, j'ai observé de cette façon la capture de plusieurs milliers d'ions, que j'ai tous comptés, et je n'en ai jamais rencontré un dont la charge, mesurée en valeur relative, comme ci-dessus, ne fût pas exactement égale à la valeur de la plus petite des charges capturées ou un très petit multiple entier de cette valeur. Voilà donc la preuve directe et incontestable que l'électron n'est pas une « moyenne statistique », mais que les charges électriques des ions sont toutes soit exactement de la même grandeur, soit égales à de petits multiples exacts de cette grandeur. »

l'électron, quantum d'électricité

Cette conclusion nous apporte le maître mot d'« électron » et nous assure que **les changements de charge se font par gain ou perte de nombres entiers d'électrons,** ces rapports entiers étant imposés par les résultats expérimentaux trouvés, sans discussion possible.

Dans la suite, Millikan montre de façon tout aussi convaincante que non seulement les changements de charge, mais aussi les charges initiales de ces gouttes contiennent un nombre entier d'électrons. Nous voici en présence de ce qu'on appelle pompeusement « l'atomisme **(quantification)** de l'électricité ».

Ici chacun posera la question : Est-ce bien sûr? Qui vous dit que cette « charge élémentaire » que vous appelez « électron » n'est pas elle-même composée de charges plus petites? Cela se peut fort bien, mais ces charges plus petites, personne ne les a jamais vues se séparer, personne n'a jamais décelé une charge électrique plus petite que celle de l'électron. Ce n'est pas faute d'avoir cherché : aux environs de 1900, certains physiciens avaient cru trouver un « sous-électron », et Millikan consacre justement un long chapitre de son livre à critiquer ces expériences et à réfuter leurs conclusions. Ses arguments sont probants, et depuis lors aucune expérience n'est venue apporter d'indications positives en faveur d'une divisibilité de l'électron.

mesure de e, *de* m *et de la constante N d'Avogadro*

Dans ce qui précède, il est question de mesures relatives des charges. La méthode de Millikan peut faire mieux et atteindre leur valeur absolue, c'est-à-dire exprimée en fonction des unités fondamentales. Ayant mesuré la charge d'une goutte et étant capable de compter le nombre d'électrons, on en déduit la valeur de la charge d'un électron, pour laquelle on trouve $1,592 . 10^{-19}$ coulomb, alors que la valeur considérée aujourd'hui comme la plus exacte est $1,602\,06 . 10^{-19}$ coulomb.

Voilà une grandeur bien petite! Il est vrai qu'en matière de charge électrique nous sommes très gourmands : le coulomb, si grand à l'échelle atomique, est une grandeur très modeste pour les applications pratiques où les quantités d'électricité s'expriment plus volontiers en ampères-heures valant 3 600 coulombs. Quand nous disons que notre batterie de voiture a une capacité de 90 Ah, cela veut dire qu'elle serait (théoriquement!) déchargée après avoir débité $90 \times 3\,600 = 324\,000$ coulombs. Laissons au lecteur le soin de compter combien cela fait d'électrons et d'exprimer le résultat en milliards de milliards.

Ces mêmes mesures fournissent une évaluation précieuse — parce qu'extrêmement directe — du nombre d'Avogadro **N,** c'est-à-dire du nombre d'atomes contenus dans l' « atome-gramme » conventionnel des chimistes. Il suffit de comparer la charge de l'électron e à la charge F (96 488 coulombs) transportée par l'électrolyse d'un atome-gramme de nitrate d'argent, pour trouver

$$\mathbf{N} = \frac{F}{e} = \frac{96\,488}{1,602 . 10^{-19}} = 6,023 . 10^{23}.$$

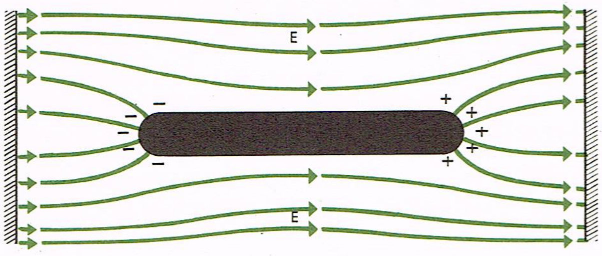
Mais nous n'avons pas encore complètement pressé le fruit de ces expériences : il était possible, au moyen de rayons X, de charger successivement une même goutte positivement et négativement. Nous ne préciserons pas ici de quelle façon, mais seul importe le résultat trouvé : les ions négatifs et les ions positifs ont des charges de même grandeur, conclusion qui résultait de la mesure des vitesses d'une même goutte que l'on faisait successivement monter et descendre dans le champ, selon que sa charge était positive ou négative.
Il est facile de vérifier que la force électrique créée par le champ est assez intense pour équilibrer le poids de la goutte. Millikan disposait d'une batterie de 5 000 à 10 000 V, alors qu'une tension de 3 000 à 4 000 V suffirait à soulever des gouttes de 1 μ (rayon moyen des gouttelettes d'huile).

la loi de conservation de l'électricité

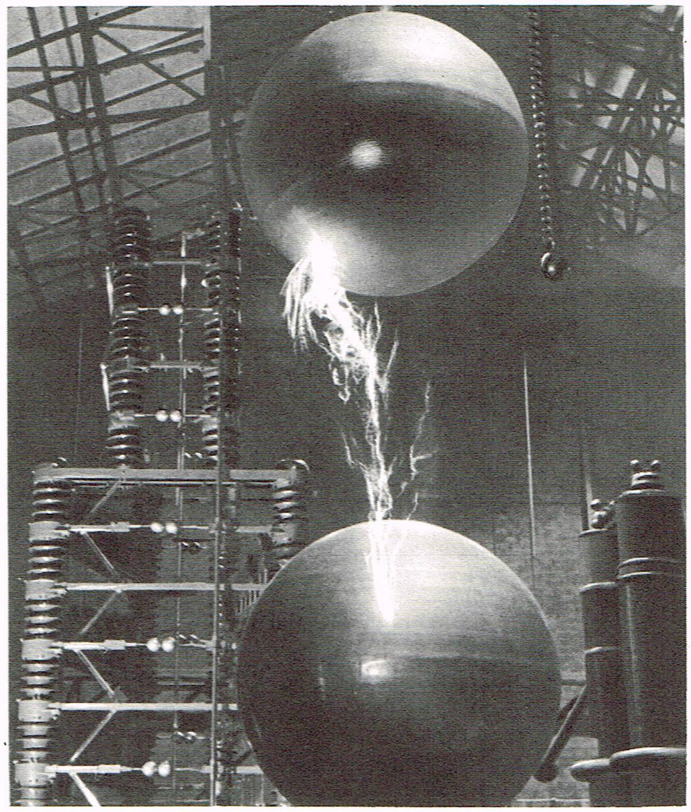
Ce résultat apporte une réponse précise à la question de la **conservation de l'électricité** qui est tout à fait primordiale. Les physiciens du XVIIIe siècle étaient déjà convaincus de la production simultanée des deux électricités de signe (on disait alors : « de nom ») contraire, en quantité égale. En témoignaient des expériences du genre de celle-ci : si deux plateaux de verre, dont l'un est recouvert de drap, sont frottés l'un contre l'autre, l'ensemble des deux disques est — et reste — électriquement neutre, tandis que si on les sépare, on constate que chacun des deux est chargé, l'un positivement, l'autre négativement. De telles expériences n'ont que la précision très limitée de l'électroscope, qui décèle l'absence de charges; on dit que ces mesures valent ce que vaut la sensibilité du zéro de l'appareil. Dans les expériences de Millikan, l'erreur dans l'évaluation des charges positives ou négatives n'atteint pas le demi-millième, ce qui est fort beau. Est-ce assez pour ériger en principe la conservation de l'électricité? Une mesure isolée, même exacte à un dix-millième près, ne suffit pas pour établir une théorie générale; en revanche, une mesure à un millième près — soit dix fois moins précise — qui vient corroborer une foule d'expériences de précision encore moindre, et qui se prête à des déductions dont les conséquences peuvent être testées presque indéfiniment, c'est un fondement solide pour une théorie.

La notion de « conservation de l'électricité » signifie donc pour nous **qu'aucune charge positive ne peut se manifester sans qu'apparaisse en même temps une charge négative égale, les deux phénomènes étant indissolublement liés.** Par exemple, lorsqu'une baguette isolante, ou un bout de fil métallique — peu importe —, est placée dans le sens des lignes de force d'un champ électrique, des charges de signe opposé se manifestent à ses deux extrémités; elles sont rigoureusement égales.

Toutefois, cette compensation peut se réaliser de bien des manières, entre lesquelles le choix ne peut se faire sans supplément d'information. On pourrait, par exemple, imaginer des particules semblables de charge égale et opposée, qui existeraient dans la matière électriquement neutre et, en venant s'attacher à des atomes ou des agrégats d'atomes neutres, en



Il n'y a pas de charges positives, seulement des charges négatives en nombre égal se créant ailleurs.



Eclateur à boules à très haute tension : la boule inférieure est reliée « à la masse »; la boule supérieure à un générateur à très haute tension (visible à l'arrière-plan). Le *champ* est si intense qu'une étincelle, foudre en miniature, éclate entre les deux boules.

Phot. R. Doisneau (Rapho).

feraient, suivant le cas, des ions positifs ou négatifs. Beaucoup de faits expérimentaux, en premier lieu les expériences de Millikan, entreraient de bonne grâce dans ce schéma.

Beaucoup d'autres faits, en revanche, nous montrent que les choses ne se passent pas ainsi; notamment ceux où se manifestent les particules électrisées lancées à grande vitesse dans le vide : les rayons cathodiques et les rayons positifs. On y voit les charges des deux signes se comporter d'une façon toute différente. Les particules négatives se montrent très légères, toutes identiques; ce sont les électrons, constituants universels de la matière. Les particules positives ont des masses beaucoup plus élevées; leur constitution ne pourra être expliquée que plus loin. Cependant, ne serait-ce que pour rendre notre langage plus clair, il faut donner dès maintenant quelques indications sommaires. L'ion positif sera regardé comme un atome ayant perdu un ou plusieurs électrons, les plus extérieurs et les moins solidement liés — ce qui n'ébranle pas gravement son architecture (il reste d'ailleurs toujours prêt à les reprendre pour redevenir atome neutre). L'ion négatif est un atome qui s'est annexé provisoirement un ou plusieurs électrons (il donne lieu aux mêmes remarques).

Chaque fois qu'un électron est libéré, il laisse derrière lui un ion positif. Qu'il vagabonde plus ou moins longtemps, plus ou moins loin, qu'il s'attache à un atome pour former un ion négatif, dans tous ces cas la conservation de la charge électrique est satisfaite. La physique corpusculaire nous fera assister à d'étranges phénomènes dans le monde des particules; pourtant, la conservation de la charge y sera tout aussi rigoureusement satisfaite.

LA PHYSIQUE DU CHAMP ÉLECTRIQUE

La démarche essentielle dans l'analyse des forces électriques consiste, avons-nous dit, dans la distinction entre ces deux éléments bien différents que sont la charge et le champ.
Nous venons de nous attarder assez longuement sur le premier de ces éléments; c'est le moment de nous intéresser au second. Dans l'expérience de Millikan, le champ était limité par deux plateaux plans et horizontaux, le « vecteur champ » avait en tout point la même direction verticale et de plus la même grandeur, tout comme le vecteur $\vec{g}$ (le principe de l'expérience étant justement de tirer parti de cette disposition pour ajouter ou soustraire la force électrique et le poids).

Un tel champ est dit « uniforme », et il est clair que c'est là une situation bien particulière : la plupart des champs électriques ne sont pas aussi simples. Ils sont — dans nos appareils, sinon toujours dans la nature — limités par un certain nombre de surfaces métalliques (électrodes) : il y en a au moins deux, reliées aux pôles d'une « machine électrique », qui peut être un générateur électrostatique, une batterie de piles ou d'accumulateurs, ou les fils d'un secteur. Dans le langage de nos électriciens, ces deux conducteurs limitant le champ sont l'un au « potentiel zéro », celui que l'on peut toucher impunément, et qui est dit encore « à la masse », l'autre à un « potentiel élevé » (par exemple + 10 000 ou — 10 000 V). Dans un champ uniforme, le produit de la grandeur du champ par la distance des plateaux s'appelle la « différence de potentiel entre ceux-ci ».

différence de potentiel et intensité de champ

Lorsqu'un champ uniforme correspond à une différence de potentiel de 1 volt entre deux plateaux distants de 1 mètre l'un de l'autre, on dit que l'intensité du champ (plus brièvement : que le champ) vaut 1 volt/mètre. (Le volt au mètre est

une unité M. K. S. A. au même titre que le volt et le mètre.) Ne faisons maintenant plus de suppositions sur la forme des électrodes. Si, en un point quelconque de l'espace les séparant, on amène (dans les expériences de Millikan c'était matériellement, désormais ce sera seulement par la pensée) une petite charge électrique ponctuelle, elle est soumise à une force dont la grandeur et la direction indiquent la grandeur et la direction du champ électrique en ce point.

lignes de force et surfaces équipotentielles

Imaginons cette action répétée en un grand nombre de points du champ, en partant d'une électrode jusqu'à l'autre : les directions du champ en ces points vont dessiner entre ces électrodes des courbes; ce sont des lignes de force, dont la forme se modifie d'une manière continue dans l'étendue du champ. Un faisceau assez serré de ces lignes dessinera une représentation très parlante du champ. Les lignes de force sont à leurs deux extrémités perpendiculaires aux électrodes.

Suivons maintenant l'une d'elles et graduons-la, non en parties d'égale longueur comme on le ferait dans un champ uniforme, mais avec des graduations d'autant plus serrées que le champ est plus intense et, par suite, d'autant plus espacées qu'il est plus faible : c'est une graduation en potentiel.

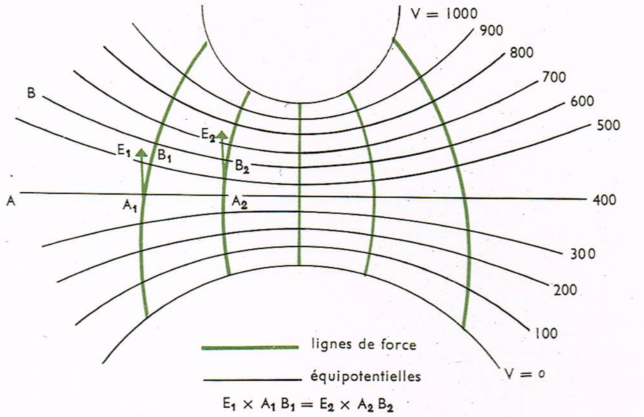
Dans l'exemple où l'une de ces électrodes est à zéro et l'autre à + 1 000 V, on pourra par exemple marquer sur les lignes de force les points variant de 100 en 100 V.

Joignons alors les points portant la même graduation sur toutes les lignes de force : nous obtenons les surfaces équipotentielles. L'une des électrodes joint tous les points marqués zéro, l'autre tous les points marqués 1 000; ce sont deux surfaces équipotentielles qui ont sur les autres l'avantage d'être non des fictions géométriques, mais des réalités bien solides. Une surface équipotentielle quelconque coupe perpendiculairement toutes les lignes de force, et on a ainsi une vraie « carte géographique » du champ, les équipotentielles et les lignes de force correspondant très exactement aux lignes de niveau et aux lignes de plus grande pente. Cette image très parlante est universellement employée, et tout le monde parle de « creux », de « vallées », de « montagnes », de « cols » de potentiel.

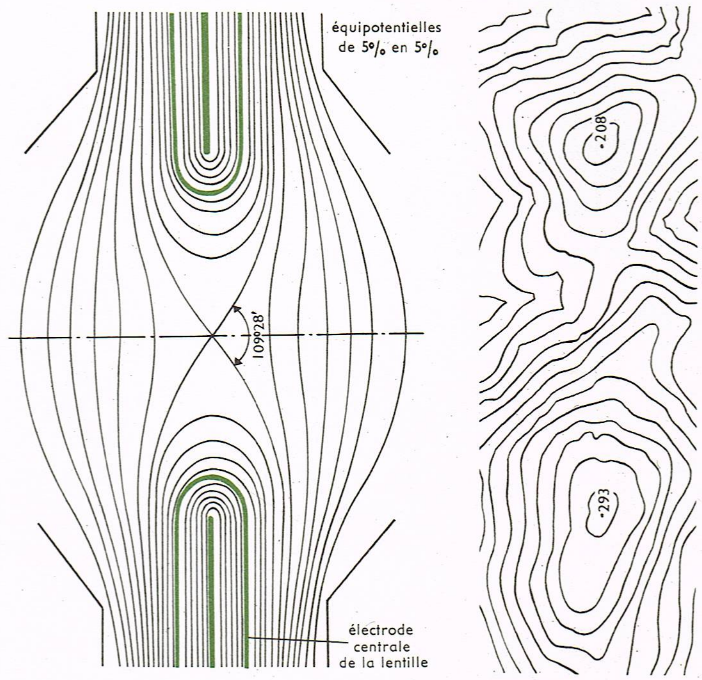
Nous avons hâte de passer de ces expériences « sur le papier » à des expériences effectives; pour cela, la nature se charge de fournir ces charges électriques idéalement petites et mobiles que sont les électrons, et la technique nous apporte le milieu convenable pour leur évolution, qui est le vide. Il s'agit maintenant, en effet, des électrons libres et non plus attachés à des gouttes liquides — minuscules à notre échelle, mais énormes à la leur —, et ils doivent se mouvoir sans se heurter constamment à l'innombrable cohue des molécules qui se bousculent dans un gaz, dès lors que sa pression est notable. C'est pour cela qu'il est si essentiel de savoir créer des vides de plus en plus poussés, et de savoir expérimenter dans leur enceinte. Et c'est ici le lieu de rappeler cette définition due à l'humour de K. K. Darrow : « La physique moderne est fille de la pompe à vide. »

matérialisation par les tubes à vide

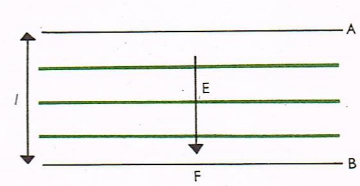
Il est tout à fait impossible de résumer, même très sommairement, plus d'un siècle d'expériences et de travaux prodigieusement touffus et variés qui ont amené, à partir des premières ébauches de Crookes, les tubes à vide à leur perfection et à leur variété actuelles. Mais nous n'avons pas ici à faire l'histoire de la physique; aussi ne retiendrons-nous que quelques exemples

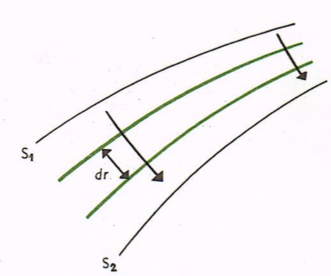


Deux électrodes et le champ électrique entre elles.



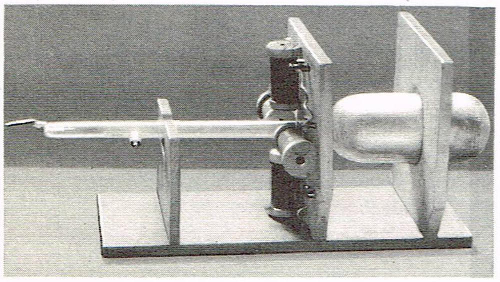
Lignes de forces et lignes de niveaux.



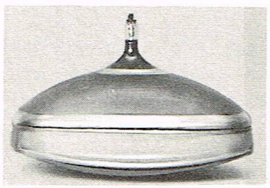


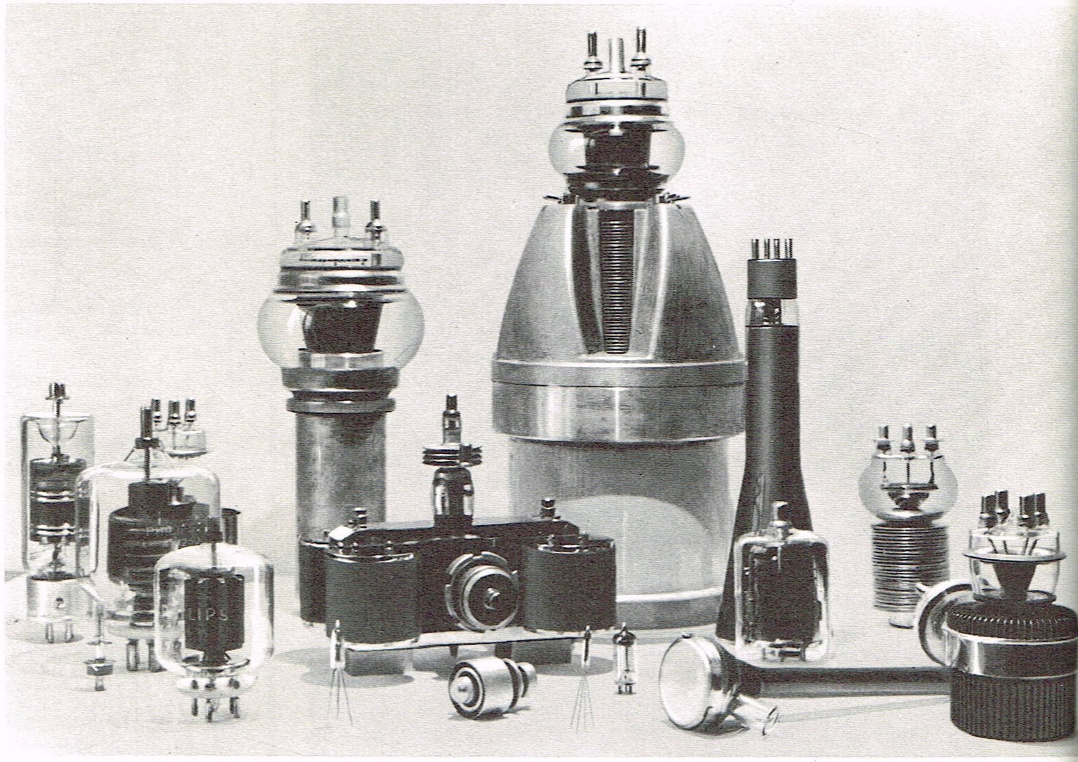
La différence de potentiel entre les surfaces A et B vaut $V = V_1 - V_2 = El$ (E intensité du champ supposé uniforme).

$dv = E dr$
La différence de potentiel entre les surfaces S_1 et S_2 vaut $V_1 - V_2 = \int dV = \int E$ (champ non constant) par définition d'une intégrale.



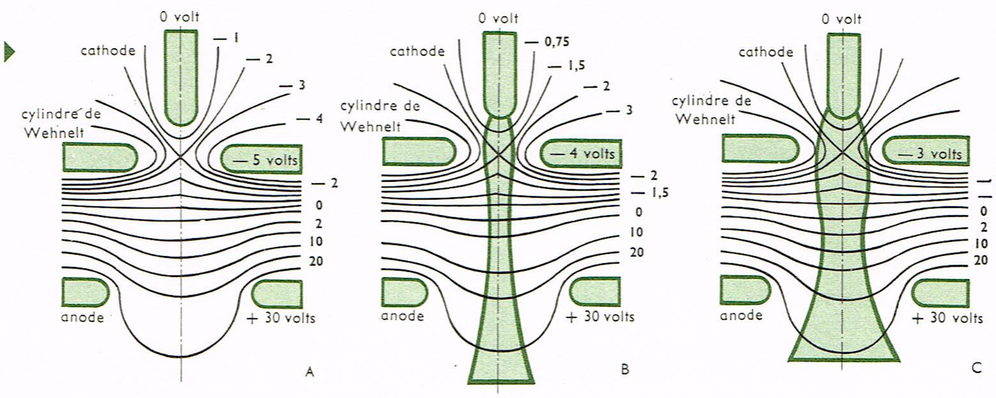
Du tube de Braun
à celui d'un téléviseur actuel.

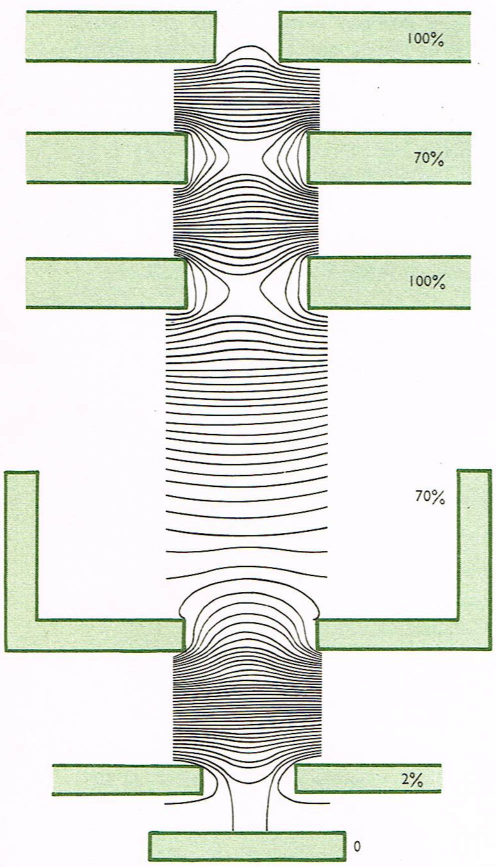




Les quelques modèles de tubes groupés ici illustrent éloquemment l'extraordinaire diversité atteinte par des dispositifs issus des premiers tubes de Crookes.

***Ci-contre,* réglage de l'intensité d'un pinceau électronique par une électrode de contrôle. (Schéma extrait de « Grundlagen der Elektronen-Optik », par W. Glaser 1952.)**





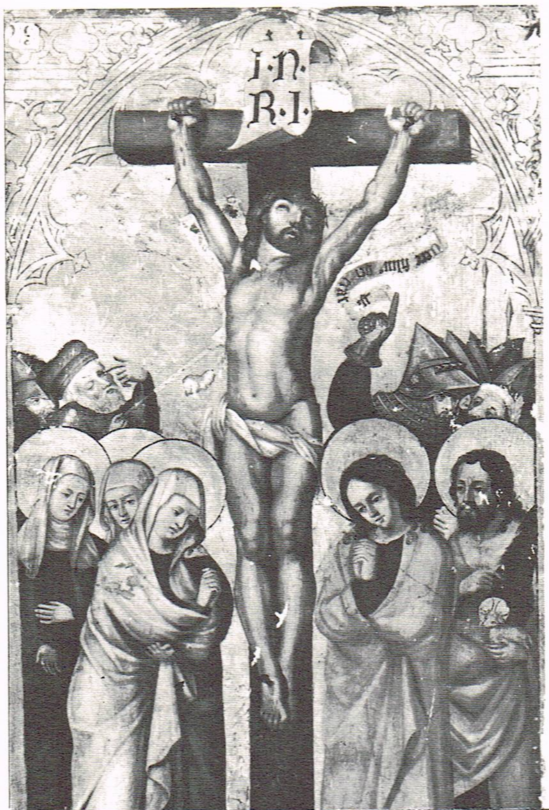
Le champ électrique dans un canon à électrons particulièrement perfectionné. (D'après W. Glaser, *op. cit.*)

modernes. La télévision a rendu familier à tous le tube à rayons cathodiques, où un pinceau d'électrons (rigoureusement invisible par lui-même, même si le tube était transparent) se manifeste par son impact sur le fond revêtu de matière fluorescente. Nous voudrions cependant prier le lecteur de détourner provisoirement son attention du tube récepteur de télévision, parce que les action subies par les électrons y sont pour une grande part dues à des forces magnétiques, dont nous parlerons plus loin, pour la reporter sur des tubes analogues employés dans les oscilloscopes, qu'un nombre croissant de non-spécialistes a vu en action (notamment les tubes récepteurs de radar) et dans lesquels seuls entrent en jeu des champs électriques.

Un tube cathodique contient de nombreuses électrodes et le champ électrique qui règne entre ces électrodes est très complexe : aussi bien ses fonctions sont-elles multiples. Un premier groupe d'électrodes situé dans la queue du tube porte le nom significatif de « canon à électrons ». Son rôle est de produire le faisceau en lui donnant les qualités désirées : intensité (c'est-à-dire nombre des électrons), énergie (vitesse

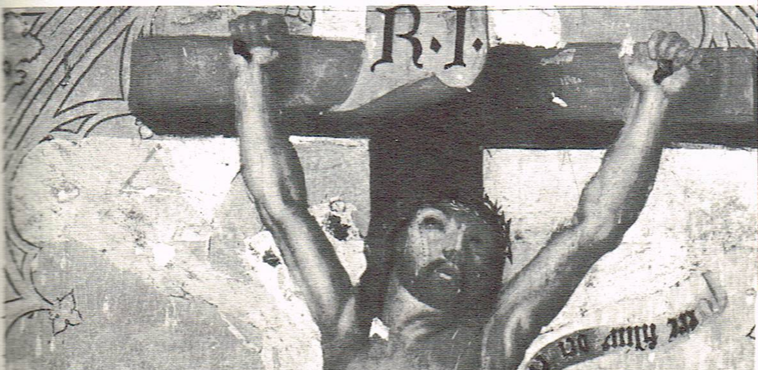
Phot. Larousse, la « Radiotechnique », Philips (Ste Hyperelec).

L'œuvre en cours de restauration (le Christ du XIVᵉ s. est encore recouvert par une peinture de l'école française du XVIIᵉ s.).

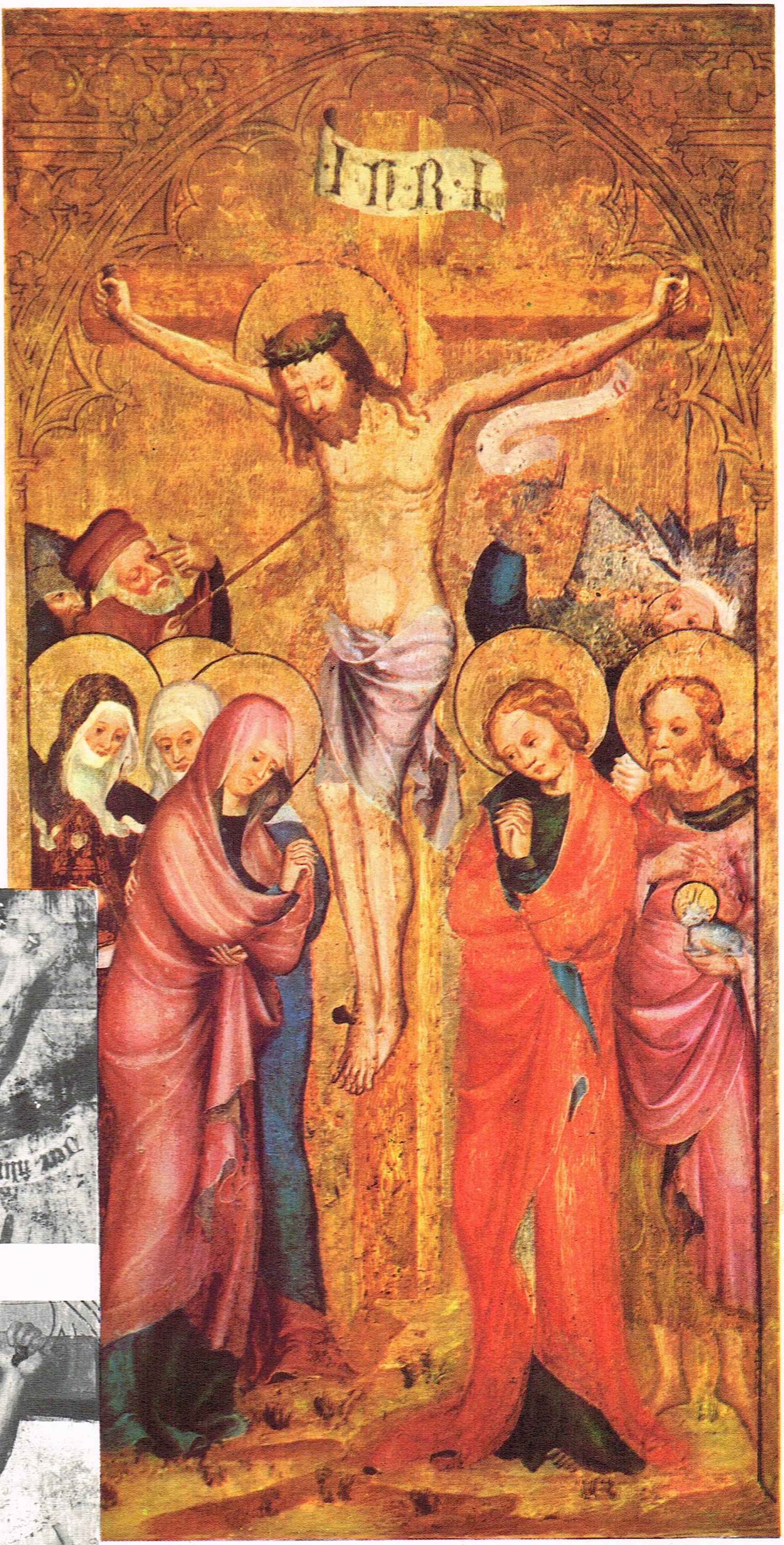




Détail montrant la superposition des deux peintures photographiées à l'ultraviolet. *Ci-dessous*, le même détail en lumière infrarouge.



Photographies du laboratoire du musée du Louvre, Institut Mainini.



« Christ en croix » (retable de la Passion : le Calvaire), *musée de Besançon, phot. Giraudon.*
Aspect actuel de la peinture due à l'école flamande ou souabe (v. 1380).

des électrons), focalisation (réunion des trajectoires sur une très petite surface de l'écran, très petite surface que nous appelons le « spot »). Dans les tubes les plus simples, le canon contient déjà quatre électrodes. La cathode (à — 3 000 volts pour fixer les idées) est la source des électrons; portée au rouge par une minuscule chaufferette électrique, elle est recouverte d'un mélange chimique complexe d'où sort un nuage d'électrons, tel un essaim de moustiques. Une partie de ces électrons sont en quelque sorte aspirés par le champ électrique et forment le faisceau. Ils doivent d'abord traverser l'électrode de contrôle (d'un potentiel de 40 à 50 V moins élevé que celui de la cathode). En réglant son potentiel, on règle le débit d'électrons, ainsi qu'on le comprend facilement en regardant le tracé des équipotentielles du champ; on y voit le col que les électrons doivent franchir : s'il est trop haut, ils ne passent plus, la valve est fermée. Les électrons qui sont passés se trouvent dans un champ accélérateur, où ils prennent de la vitesse. Ce champ est créé par les deux anodes, respectivement à — 1 000 V et à zéro (il reviendrait au même de dire que la cathode est à zéro, l'électrode de contrôle à — 50, la première anode à + 1 000, la seconde à + 3 000 : l'origine du potentiel dans un champ, comme le niveau de départ des altitudes d'un pays, peut être choisie arbitrairement). Pourquoi deux anodes, quand une seule suffirait à donner de la vitesse aux électrons? C'est que la deuxième réalise un champ moins simple, et, partant, plus riche de possibilités, pour infléchir les trajectoires et les amener à se concentrer sur l'écran.

l'optique électronique ou le calcul de difficiles trajectoires

Ces trajectoires ne sont pas si faciles à prévoir. N'oublions pas que la direction du champ en chaque point nous indique l'accélération et nullement la vitesse de l'électron qui y passe; aussi le tracé des trajectoires complètes demande-t-il des calculs considérables. Le calcul des trajectoires et la synthèse des champs qu'il faut réaliser pour avoir celles que l'on souhaite — opération en quelque sorte inverse de la première — font l'objet d'une discipline particulière, intéressante et difficile : l'optique électronique. Son nom indique assez sa nature hybride entre l'imagerie optique (trajectoires correspondant aux rayons lumineux) et la physique de l'électron. Ce dernier consent, dans la plupart des problèmes posés par les tubes, à se comporter comme un « honnête » point matériel gouverné par les lois de la dynamique.

Avant d'aller plus loin, il faut compléter notre description du tube cathodique. Limité à ce qui précède, il enverrait le pinceau électronique frapper l'écran toujours au même point, où il n'aurait d'autre occupation que de faire fondre la paroi, effet désagréable pour tous. Pour que le pinceau vienne peindre sur l'écran, il faut le mouvoir; cela est l'affaire des champs déviateurs.

Notre propos actuel est de faire un peu mieux connaissance avec l'électron. « Honnête point matériel », avons-nous dit. « Honnête », on en reparlera; « matériel », à vrai dire si peu! La masse est infime et sa charge si grande en comparaison que la moindre pichenette lui communique des vitesses très élevées. Nous connaissons déjà sa charge; quant à la masse, c'est seulement un peu plus loin que nous aurons les éléments nécessaires pour expliquer le principe de sa mesure. Nous nous contenterons donc provisoirement d'enregistrer le résultat : elle vaut $0{,}910\,83 \cdot 10^{-27}$ gramme.

La mécanique nous a appris qu'une masse m soumise à une différence de potentiel H acquiert une énergie mH. De même, une particule, initialement immobile, de charge e soumise à une différence de potentiel V acquiert l'énergie eV liée à la vitesse V par la relation $eV = \frac{1}{2} mv^2$.

A un électron ($e/m = 176 \cdot 10^9$), une différence de potentiel de 100 volts communique une vitesse de 6 000 km/s, et une différence de 10 000 volts (cent fois plus grande) une vitesse de 60 000 km/s (dix fois plus grande).

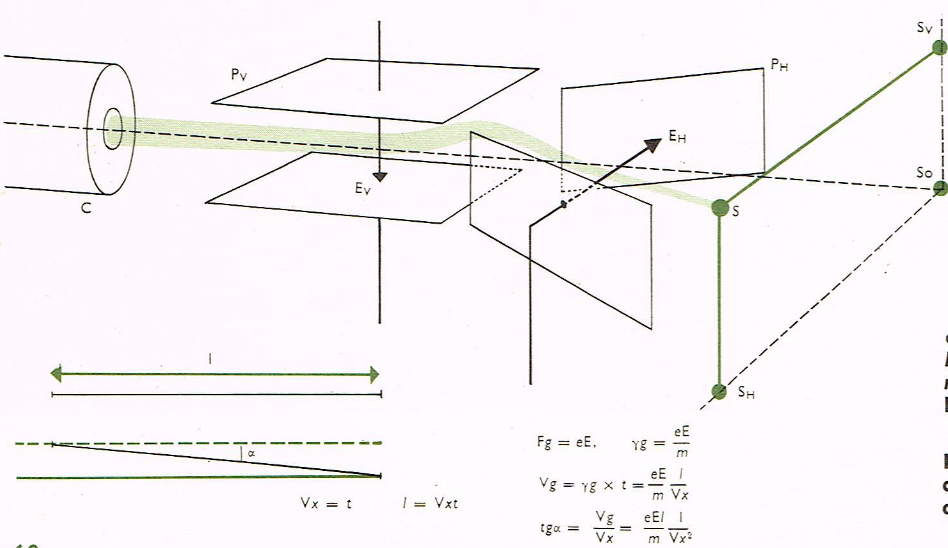
Deux particules de même charge e, mais de masses respectivement m et m', acquièrent des vitesses v et v' différentes : $\frac{v'^2}{v^2} = \frac{m}{m'}$ (produit mv^2 constant).

A un proton (atome d'hydrogène privé de son unique électron) de même charge, mais de masse 1 800 fois plus grande, une même différence de potentiel communique une vitesse 45 fois ($\sqrt{1\,800}$) plus faible qu'à un électron (135 km/s et 1 333 km/s dans l'exemple numérique indiqué).

mouvement de charges et courant

Le courant électrique tel que nous le manipulons tous les jours dans nos circuits d'éclairage ou de radio est un transport de charge. Par définition même, un courant de 1 ampère transporte 1 coulomb en 1 seconde (unité M. K. S. A.).

Par suite, un pinceau d'électrons en mouvement dans le vide se comporte comme un courant. De quelle intensité? La figure ci-dessous montre que N électrons de charge unitaire e se déplaçant à la vitesse v correspondant à un courant Nev (remarquons

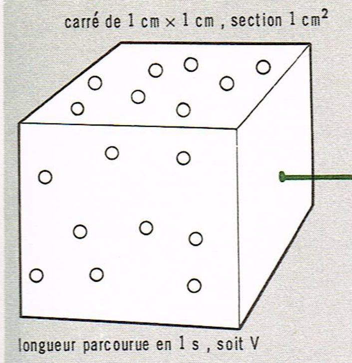


Déviation du pinceau électronique

C canon à électrons
Pv, Ev plaques et champ déviateur verticaux
PH, EH plaques et champ déviateur horizontaux
So spot en l'absence de champ déviateur
SH spot sous l'action du champ EH nul
Sv spot sous l'action du champ Ev nul
S spot sous l'action combinée EH et Ev

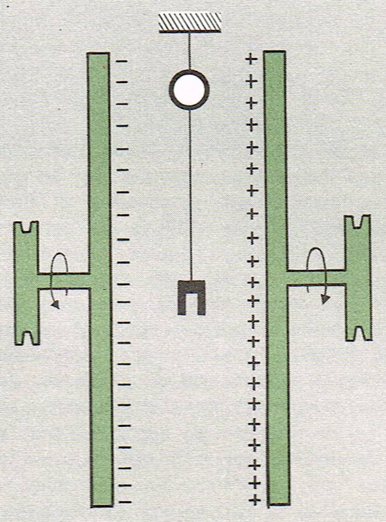
e charge de l'électron
l longueur du tube déviateur
m masse de l'électron (*e/m* = ...)
E intensité du champ déviateur

La déviation varie comme l'inverse du carré de la vitesse des électrons (e, E, *l*, *m* sont des constantes universelles ou instrumentales).

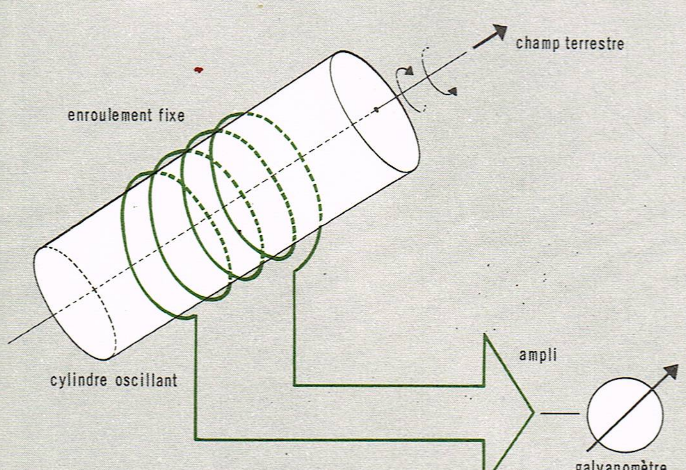


Chaque charge vaut e; il y en a *n* par cm³ : le volume vaut v cm³, la charge totale *nev*. Elle traverse en 1 s une section de 1 cm² : c'est une *densité de courant*. Si N est le nombre total d'électrons dans la section droite du pinceau, Nev est le courant total transporté par celui-ci.

Expérience de Rowland.



Deux disques aux charges contraires tournent en sens inverse pour doubler l'effet. Au centre, petit aimant suspendu et miroir pour déceler le champ magnétique.



Expérience de Tolman.

L'axe du cylindre oscillant est dirigé suivant le champ magnétique terrestre pour prévenir les effets d'induction autres que ceux qui sont recherchés.

que la charge e de l'électron étant négative, le sens conventionnel du courant est opposé à celui de *v'*), le courant Nev vaudra un milliampère pour $N = \frac{0{,}001}{(6.10^7)\,(1{,}6.10^{-19})} = 10^8$ (v = 60 000 km/s ou 6.10^7 m/s; — e = $1{,}6.10^{-19}$). On trouve donc qu'à ce courant correspond le déplacement de 10^8 (cent millions) électrons.

Les physiciens n'ont pas attendu l'avènement des tubes à vide pour chercher la preuve directe que le mouvement de charges électriques est réellement un courant. On a chargé d'électricité « statique » un disque métallique et on l'a fait tourner rapidement dans son plan, en cherchant si cela produirait un champ magnétique sur l'axe comme le ferait un courant circulaire. Non sans peine (expériences de Rowland en 1876, reprises par Cremieu et Pender en 1903), on est parvenu à un résultat certain et positif, bien que l'effet soit très petit (le champ magnétique qu'il s'agit de déceler est de l'ordre d'un dix-millième du champ terrestre). Les vitesses qu'on peut donner au disque matériel (sans risque de le voir éclater sous l'effet de la force centrifuge) sont insignifiantes en comparaison des vitesses qu'atteignent les électrons tout seuls quand on les met en branle par un champ électrique en laissant immobile, avec son énorme inertie, le métal qui les contient. C'est pour faire ressortir ce contraste entre la légèreté de l'électron et la pesanteur de la matière que nous avons cité cette vieille expérience. En voici maintenant une autre, qui montre que cette légèreté de l'électron n'est pas illimitée : un cylindre creux en laiton oscille autour de son axe; si les électrons qu'il contient étaient totalement dépourvus d'inertie, ils seraient entraînés entièrement. Au contraire, comme ils en possèdent une, l'entraînement n'est que partiel, et leur déplacement relatif se traduit par la naissance d'un courant (alternatif). D'ailleurs, un tel courant, même très faible, peut être décelé par les courants qu'il induit dans une bobine fixe. Ces expériences (Tolman et ses élèves, 1912-1925) demandent une grande virtuosité, tant les phénomènes à mesurer sont petits; elles sont très frappantes, parce qu'elles font toucher du doigt l'identité entre le courant et le mouvement des électrons, avec l'inertie mécanique de ces derniers. Le calcul montre qu'on peut ainsi mesurer le rapport e/m; détermination d'ailleurs peu précise, mais qui recoupe parfaitement celles, moins directes, que nous verrons plus loin.

l'électron, prisonnier de la matière

En regard de l'incomparable liberté de mouvement des électrons dans le vide, la situation est très différente dans les métaux et, plus généralement, dans les solides conducteurs. Les électrons y restent pratiquement libres pour des déplacements de très faible amplitude (comme dans les expériences auxquelles nous venons de faire allusion), mais leurs migrations à moyenne ou longue distance, qui constituent la forme la plus habituelle du courant électrique, se font beaucoup moins allégrement.

Ce problème constitue une partie essentielle de la physique des solides; contentons-nous ici d'une très sommaire introduction, en attendant l'étude plus complète qui en sera faite plus loin.

Il n'est pas facile de se représenter le mouvement d'un électron dans un métal, et d'abord la structure du métal « vue » par un électron. Les atomes y sont rangés en un réseau et s'alignent en files régulières répétant indéfiniment le même motif. Pour ceux qui ont eu le privilège de visiter l'inoubliable cathédrale de Cordoue, ses colonnades, qui, d'un point quelconque de l'édifice, paraissent dans la demi-obscurité se prolonger indéfiniment en tous sens, en donnent si l'on veut une image; une plantation bien régulière de peupliers en donnerait une autre à moins de frais. En revanche, les électrons de chaque cristal du métal forment une foule dense qui s'agite et se bouscule dans toutes les directions : ils se heurtent les uns aux autres et aux atomes. Qu'un champ électrique soit établi dans le

cristal, et les électrons vont lui obéir, du moins dans l'intervalle des chocs. Leur trajectoire reste irrégulière, mais comporte alors une dérive moyenne dans la direction du champ : c'est ce qui s'appelle la conduction. La mesure du courant est celle du débit des électrons, c'est-à-dire du nombre d'entre eux qui traversent en une seconde la section droite du conducteur. Une grandeur un peu plus abstraite, indispensable à l'intelligence des phénomènes, est celle de la densité de courant, qui indique à la fois le débit par unité de section et la direction dans laquelle se fait l'écoulement; elle a donc comme le champ $\vec{E}$ les caractères d'un vecteur, et ce vecteur densité de courant jouera un rôle essentiel dans la genèse des ondes électromagnétiques.

l'art et la manière de secouer les électrons, mesure du progrès technique

Ce qui précède fait ressortir la bonne et la mauvaise manière de « secouer les électrons », pour reprendre une pittoresque expression due à Charles Fabry. La manière efficace est celle qui s'adresse à leur charge, par l'action du champ électrique; la manière inefficace est celle qui s'adresse à leur masse, en essayant de leur imposer un mouvement de corps matériel. L'important est, en effet, de bien savoir agiter les électrons : leur écoulement chaotique dans un fil métallique le porte à l'incandescence; leur assaut contre les atomes ou les molécules des gaz provoque l'émission de la lumière propre à chacune des particules excitées; leur mouvement rythmé dans une antenne est responsable de l'émission radio. Mais combien complexes sont ces mouvements, et combien difficilement contrôlables : l'emprise que nous avons sur eux sert en quelque sorte de mesure au progrès technique.

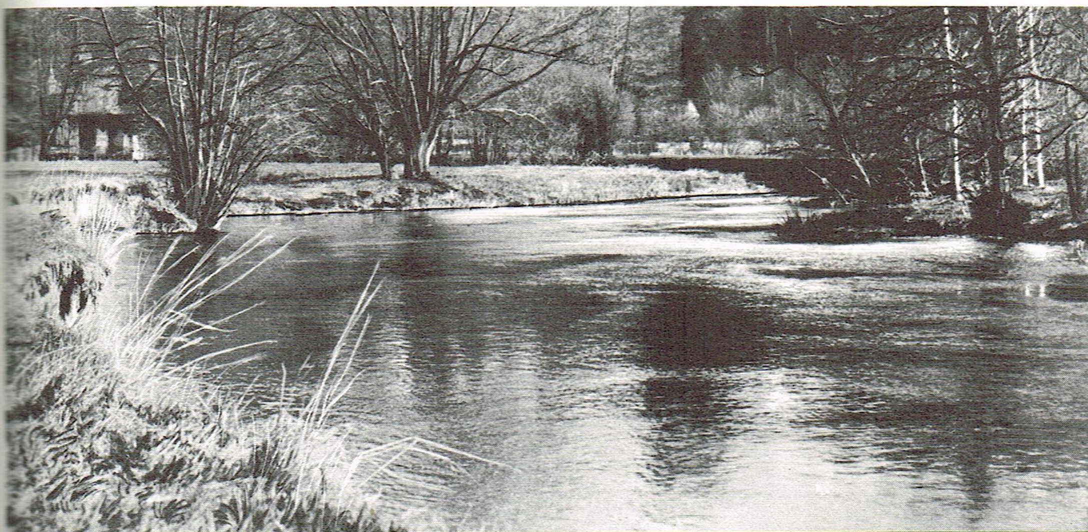
Il a déjà été fait allusion à l'agitation spontanée, désordonnée, perpétuelle dont les électrons sont animés au sein des métaux (plus généralement des solides), de plus en plus vive à mesure que la température s'élève et qui ne disparaît qu'au zéro absolu. Phénomène mystérieux, que l'on comprendra peut-être mieux après la lecture du chapitre des probabilités, et qui pose au physicien comme au philosophe des problèmes jusqu'ici non résolus. Phénomène essentiel, responsable du *bruit de fond* dont l'importance apparaîtra en pleine lumière dans le chapitre relatif à la théorie de l'information.

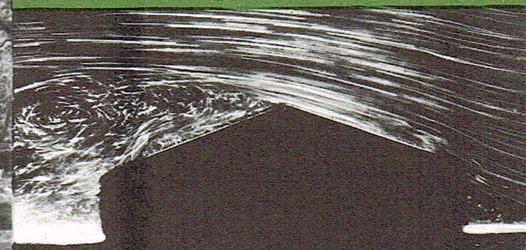
C'est enfin le moment de signaler une autre classe de mouvements des électrons, beaucoup plus mystérieux encore : les mouvements qu'ils ont à l'intérieur même des atomes. Pour une part ces mouvements sont justiciables des lois de la mécanique; ils sont produits par des champs électriques et il est certain que des champs intenses règnent à l'intérieur de l'atome et surtout au voisinage de son noyau. La mécanique nous indique alors que les électrons parcourent des orbites autour du noyau comme les planètes autour du Soleil. Mais que l'on tire les conséquences de cette comparaison pour les confronter avec les faits (révélés notamment par la spectroscopie), et les difficultés apparaissent, grossissent et deviennent vite insurmontables. Une refonte radicale de toutes nos idées, par la **mécanique quantique,** s'impose. C'est dans ce sens que l'électron cesse d'être un « honnête » point matériel; en fait, plus l'on scrute son comportement, plus on y découvre de choses étonnantes, celles que nous révèlent les chapitres de la physique corpusculaire. Nous pouvons alors pénétrer dans ce qui est le jardin le plus secret de l'univers, du moins en l'état actuel de nos connaissances scientifiques.

Par leur alignement et leur répétition, les colonnades de la cathédrale de Cordoue peuvent donner une image de la représentation de la structure d'un métal.

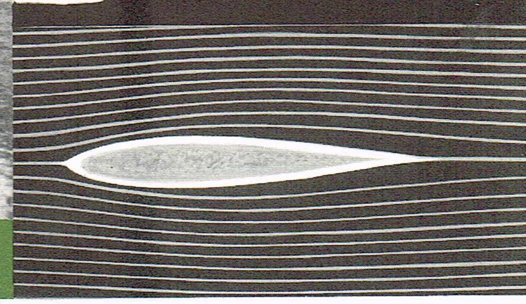


Phot. Anderson, Giraudon.





Écoulement rendu visible par de la poudre d'aluminium.



Écoulement d'une rivière au cours paisible.

Écoulement matérialisé par des filets de solution colorée.

LES CHAMPS HYDRODYNAMIQUES

Pour bien faire comprendre comment les phénomènes électromagnétiques s'enchaînent les uns aux autres, un détour apparent est nécessaire, mais qui constituera en réalité le chemin le plus direct. Bien des propriétés du champ électrique nous échappent encore, et nous aurons encore plus besoin de secours lorsque nous aurons à parler du champ magnétique. Bien entendu, l'analyse mathématique est capable de les expliquer, mais sous une forme abstraite. C'est aux écoulements de liquide que nous demanderons, au contraire, des modèles concrets (relativement concrets).

champ et flux mécaniques

Drôle d'idée apparemment, à la vue d'une rivière tranquille, mais non paresseuse, de l'appeler **champ de vitesses :** toujours cette terrible manie de l'abstraction. Mais que l'on nous pardonne : il n'y a pas d'autre moyen pour la connaissance scientifique. Jouons donc le jeu et regardons deux photographies d'écoulement. La première représente un liquide dans lequel on a mis en suspension une poudre métallique; chaque grain entraîné dans l'écoulement est éclairé pendant la durée de la pose et dessine un petit segment de droite qui représente son vecteur vitesse. C'est la vitesse du petit volume liquide qui entoure la particule (très petit volume pour la nature — infiniment petit pour la pensée). L'ensemble de ces segments concrétise exactement le champ de vitesses.

Le seconde photographie est obtenue en injectant dans de l'eau une solution colorée, qui s'y étire en filaments sans s'y mélanger sensiblement; voilà bien les trajectoires de petits volumes liquides, donc des lignes de force du champ de vitesses, qui matérialisent exactement la définition que nous en avons donnée à propos du champ électrique.

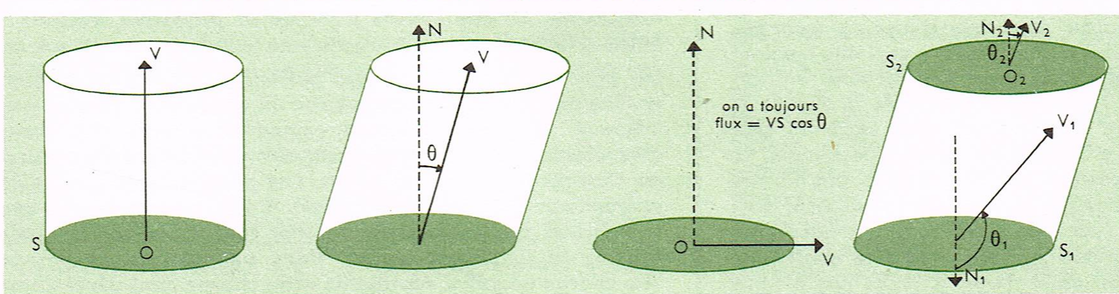
Plongeons un cerceau dans le courant; il est facile, sinon de mesurer effectivement, du moins de se représenter la quantité d'eau qui traverse le cerceau en une seconde : c'est le *flux* du champ à travers la surface du cerceau. Examinons le problème, d'un peu plus près. Comme il n'est pas question d'isoler, de soutirer cette quantité d'eau sous peine de détériorer gravement l'écoulement, c'est un calcul qu'il nous faut effectuer. D'abord, supposons que la vitesse soit la même sur toute la surface du cerceau; alors c'est bien simple : en une seconde, chaque point du liquide avance d'une longueur égale à sa vitesse et la quantité qui a traversé emplit un cylindre qui a le cerceau pour base et le vecteur vitesse pour génératrice. Pour une longueur donnée de celle-ci, le volume est le plus grand possible quand le cylindre est droit, de plus en plus petit quand il s'aplatit, nul quand il est tout à fait plat, c'est-à-dire quand la vitesse est contenue dans le plan du cerceau.

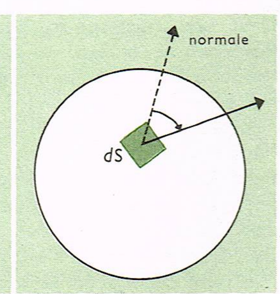
Et maintenant, qu'advient-il si la vitesse est différente d'un point à un autre de la surface du cerceau? Il faut découper

Les normales N_1 et N_2 sont orientées *vers l'extérieur* de la surface formée : par suite, θ_1 est obtus et θ_2 aigu. Le flux $\Phi_1 = S_1 V_1 \cos\theta_1$ est donc négatif; $\Phi_2 = S_2 V_2 \cos\theta_2$ est positif et l'on a $\Phi_1 + \Phi_2 = 0$. Si l'on prend l'ensemble des lignes de force s'appuyant sur une courbe fermée tracée dans un champ, il forme une sorte de cylindre nommé « tube de force » : on les utilise souvent en électromagnétisme. La surface fermée que l'on vient de considérer est une tranche d'un tube de force.

Flux à travers une surface quelconque

$$\Phi = \iint_{\text{surface totale}} V \cos\theta \, dS$$





Phot. Gris.

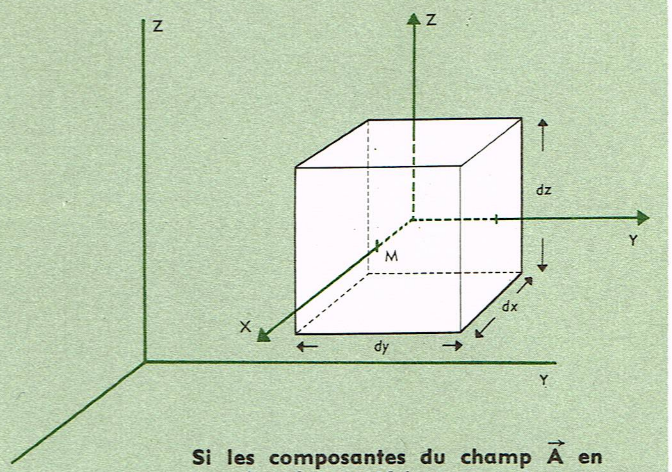
celui-ci en morceaux assez petits pour qu'en chacun d'eux la vitesse puisse être regardée comme constante, construire sur chacun d'eux un cylindre comme il vient d'être fait et additionner les volumes. Combien de morceaux? Si les vitesses ne sont pas trop différentes, et si je ne suis pas trop exigeant sur leur constance en chaque portion de surface, un petit nombre suffira; à mesure qu'une plus grande précision sera nécessaire, il faudra diviser la surface en un plus grand nombre d'éléments. Quand on voudra une représentation absolument, « intégralement » fidèle de la quantité recherchée, il faudra imaginer une infinité d'éléments à additionner. Il faudra procéder à ce que les mathématiciens appellent justement le « calcul intégral », et le résultat de ces calculs sera une « intégrale de surface ». Le calcul intégral est un art difficile, mais — comme l'art de la guerre — tout d'application. Son principe, lui, est d'une grande simplicité.

Mais il ne faut pas s'en tenir au flux qui traverse une surface plane. Pour définir le flux qui traverse une surface courbe, nous procéderons, exactement de la même façon que tout à l'heure, à sa fragmentation en un assez grand nombre de morceaux pour que chacun d'eux puisse être regardé comme plan. Nous laissons au lecteur le soin de poursuivre le raisonnement dans ses détails. Enfin, au lieu d'une surface plane ou courbe s'appuyant sur un contour, imaginons une boîte, par exemple un cube en fil de fer, à travers toutes les faces duquel le liquide coule librement. Faut-il d'ailleurs nous limiter à des boîtes formées de faces planes? Nullement, et dans la suite il sera question de flux à travers une surface quelconque, notamment à travers une surface fermée quelconque, sans qu'on ait à en préciser la forme.

Au sujet de cet écoulement à travers une surface fermée, que nous dit le sens commun? Qu'il sort exactement autant d'eau par certaines parties de la boîte qu'il en est entré par d'autres faces. Que nous dit le calcul du flux? Qu'il faut, si l'on veut distinguer l'entrée et la sortie du liquide, donner au flux à travers chaque portion de surface un signe positif ou négatif et faire ensuite l'addition algébrique des flux traversant les différentes portions. C'est particulièrement simple à concevoir lorsque la surface considérée a la forme d'un prisme et que les vecteurs vitesses sont parallèles à sa surface latérale. Le flux à travers toute cette partie est nul, et il ne reste qu'à considérer les flux à travers les deux bases; l'un étant négatif et l'autre positif. Pour un écoulement d'eau, cette somme algébrique est toujours nulle. Il pourrait ne pas en être de même dans l'écoulement d'un fluide compressible, car alors on pourrait avoir, au moins temporairement, accumulation ou au contraire raréfaction à l'intérieur de la boîte.

sources ou pertes du champ

Revenons à un écoulement d'eau; plutôt qu'une rivière, imaginons maintenant un bassin alimenté par un tuyau immergé et s'écoulant par un trop-plein, et posons-nous quelques questions sur les flux de la vitesse à travers les surfaces fermées (géométriques et non matérielles) tracées à notre fantaisie au sein du liquide. Si la surface ne comprend ni la bouche du tuyau d'arrivée ni celle du trop-plein, le flux total sortant est nul; si elle contient la bouche du tuyau d'arrivée (source), le flux sortant est égal au débit d'alimentation, et si elle contient la bouche du trop-plein (la perte), c'est le flux total entrant qui a cette même valeur. En somme, si l'on évite la source et les pertes, et que l'on trace autour de chaque point au sein du liquide une petite surface fermée de forme quelconque, le flux sortant de cette petite surface est nul. N'est-ce pas là une propriété locale bien caractéristique (la surface pouvant être tracée aussi petite que l'on veut autour du point qui nous intéresse). Elle doit donc s'exprimer par une propriété analytique du vecteur champ en chaque point. Nous n'avons pas à faire



Si les composantes du champ $\vec{A}$ en coordonnées cartésiennes sont appelées X, Y, Z (chacune d'elles est fonction des 3 coordonnées *xyz* du point M considéré du champ), l'expression de la *divergence* est

$$\operatorname{div} \vec{A} = \frac{\delta X}{\delta x} + \frac{\delta Y}{\delta y} + \frac{\delta Z}{\delta z}.$$

Il est facile de montrer que le flux sortant d'un petit parallélépipède centré sur M et de côtés *dx, dy, dz* est égal au produit de la divergence du champ en M par son volume :

$$\varphi = \operatorname{div} A.\ dx\ dy\ dz.$$

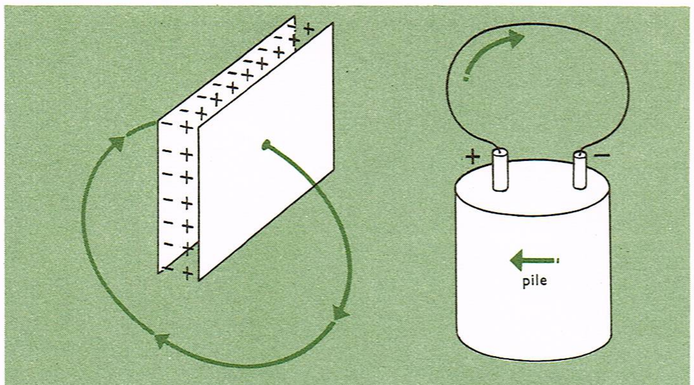
de mathématiques; ce qui nous intéresse actuellement, ce n'est pas l'expression, mais seulement l'existence de cette propriété : au lieu donc de dire que dans un écoulement de liquide incompressible le flux sortant de petites surfaces tracées arbitrairement autour de chaque point est nul, nous dirons, ce qui est plus bref, que **la divergence du champ de vitesse est nulle partout où il n'y a ni source ni perte.**

champ et flux électriques

Et maintenant retournons au champ électrique. En quoi ce qui précède le concerne-t-il? Peut-on définir le flux et comment? Un instant de réflexion montre que la définition est valable sans changer un mot. Mais qu'est-ce qui coule? Rien évidemment, et il serait fallacieux de rechercher une représentation tangible du flux : il n'est qu'une notion abstraite, analytique, mais qui permettra d'exprimer dans un langage plus commode les propriétés de tels champs et d'unifier des phénomènes physiquement très différents.

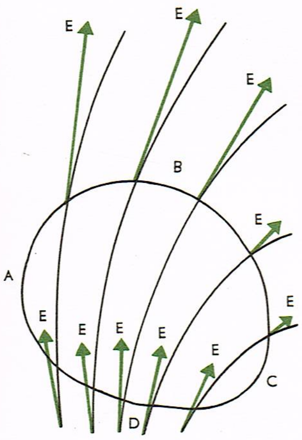
Cela peut paraître un mince avantage; c'est en réalité beaucoup, et une bonne part de la physique découle de cette méthode.

Les lignes de force électriques partent d'un conducteur chargé positivement; elles aboutissent sur un conducteur chargé négativement. Si nous faisons la comparaison avec des champs d'écoulement, les charges positives représentent les sources, les charges négatives les pertes. Les propriétés du flux se correspondent; le flux total sortant d'une surface enveloppant une charge positive représente cette charge considérée comme source; le flux total entrant dans une surface enveloppant une charge négative représente cette charge considérée comme



Courant temporaire : *sources* **sur la plaque chargée positivement,** *pertes* **sur la plaque chargée négativement.**

Courant permanent : il passe aussi bien dans la pile que dans le fil extérieur : ni source ni perte.



L'intégrale le long de la courbe ABCDA

$$\int E dr = \int_{AC} E dr + \int_{CA} E dr = (V_C - V_A) + (V_A - V_C) = 0$$

est nulle.
La circulation du vecteur le long d'une ligne fermée est nulle.

perte[1]. Nous disons ici « représente » et non plus « est égal à », car tout de même la transposition a des limites, et si l'on veut que les mesures s'agencent convenablement, on dira que le flux et la charge sont proportionnels, le rapport entre la charge et le flux étant une constante, aussi longtemps du moins que le champ électrique occupe un milieu physique homogène. La constante relative au vide, le plus naturel de tous les milieux, joue un rôle particulièrement important et s'appelle ε_0 (la constante k de la loi de Coulomb vaut précisément $1/4 \pi \varepsilon_0$). A l'extérieur de toute charge, la divergence du champ électrique est nulle. Mais que dira-t-on du champ électrique créé par des charges réparties un peu partout, par un nuage d'électrons, par exemple? Ici l'analogie hydrodynamique s'estompe : mais ce n'est pas une raison pour abandonner les notions qui en ont été tirées. Disons plutôt qu'un tel champ a une divergence différente de zéro, et proportionnelle en chaque point à la densité de charge présente.

le champ de densité d'un courant

Voici maintenant un champ plus proche d'un écoulement liquide : c'est le champ de **densité de courant.** Ici, nous savons reconnaître notre vieille connaissance, le flux : c'est tout simplement la charge qui s'écoule à travers la surface considérée; c'est le flux d'électrons au sens littéral. « Eh! que ne le disiez-vous plus tôt! » pourrait-on alors nous reprocher. **L'intensité du courant est donc tout simplement le flux de la densité de courant à travers la section du conducteur.**

Nous noterons cependant une singulière différence entre deux sortes de courants : le courant temporaire, qui va d'une armature à l'autre d'un condensateur préalablement chargé, qui diminue avec le temps et qui disparaît dès que toutes les charges se sont écoulées — ce courant a ses sources sur l'armature positive, et ses pertes sur la négative —; et, d'autre part, le courant permanent, qu'une pile produit dans un circuit — ses lignes de courant sont fermées, elles n'ont ni source ni perte. La divergence de la densité de courant est partout nulle dans ce dernier cas; elle ne saurait donc l'être dans le premier, qui comporte perte et source. Le progrès en physique consistant à découvrir des lois communes sous des apparences diverses, nous aurons à revenir sur cette question et à montrer comment une hypothèse due à Maxwell généralise la notion de densité de courant et opère la synthèse désirée.

retour aux champs hydrodynamiques

Nous venons de concrétiser par les écoulements fluides l'idée de flux, avec toutes les conséquences qui en découlent pour en faire profiter le champ électrique. Celui-ci va immédiatement payer sa dette, si l'on peut dire, en mettant à la disposition de l'hydrodynamique la notion de potentiel. Nous avons appris à tracer dans un champ des surfaces équipotentielles, partout normales aux lignes de force, et nous savons qu'entre deux équipotentielles voisines, le produit du champ par la distance est constant, quelle que soit la ligne de force sur laquelle ils sont mesurés.

Le champ de vecteurs possède donc une propriété tout à fait remarquable : **la circulation du vecteur le long d'une ligne fermée est nulle,** comme le montre la figure.

Cette propriété, le champ électrostatique la partage avec celui des liquides incompressibles s'écoulant simplement d'une source à une perte.

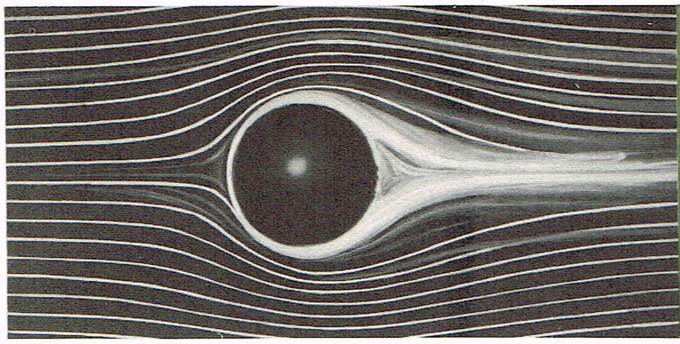
potentiel et tourbillons

En est-il de même pour tous les écoulements liquides?
Revenons donc au bord de la rivière et regardons le courant en aval d'un obstacle: Une simple branche suffit à produire de nombreux tourbillons. Dans chacun d'eux, le liquide est en rotation autour d'un axe à peu près vertical qui viendrait percer la surface au fond du petit entonnoir manifestant le tourbillon à nos yeux. Il n'y a qu'une simple différence d'échelle entre ce discret phénomène et les tourbillons de Charybde ou l'effrayant Maelström (surtout si nous considérons ce dernier amplifié par l'imagination d'Edgar Poe et à travers ses « histoires extraordinaires »). S'ils n'ont rien qui puisse effrayer, les tourbillons de la rivière doivent faire réfléchir : comment ne pas voir qu'ils bousculent l'image du champ d'écoulement que nous avons tracée? Il faut y regarder de plus près.

Nous pouvons toujours imaginer, définir et même matérialiser les lignes de courant. Certaines d'entre elles sont cependant insolites; elles sont vaguement circulaires, du moins en courbe fermée, et entourent l'axe du tourbillon. Délimitons un petit volume par une surface fermée de manière à éviter soigneusement l'axe du tourbillon; on peut s'en approcher, mais défense d'y toucher. L'écoulement est tout à fait normal (c'est-à-dire conforme au déjà vu) : flux bien défini; flux total nul d'ailleurs, définition commode des surfaces équipotentielles. Ces dernières

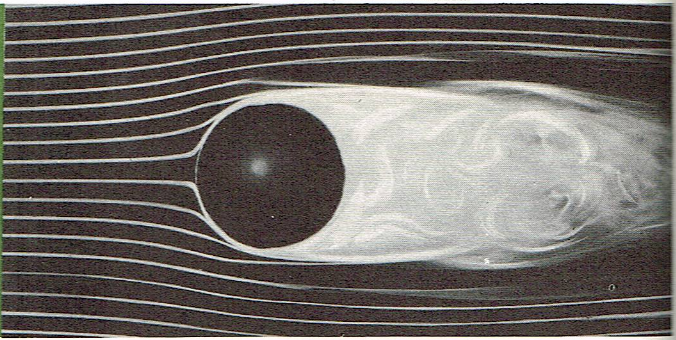
---

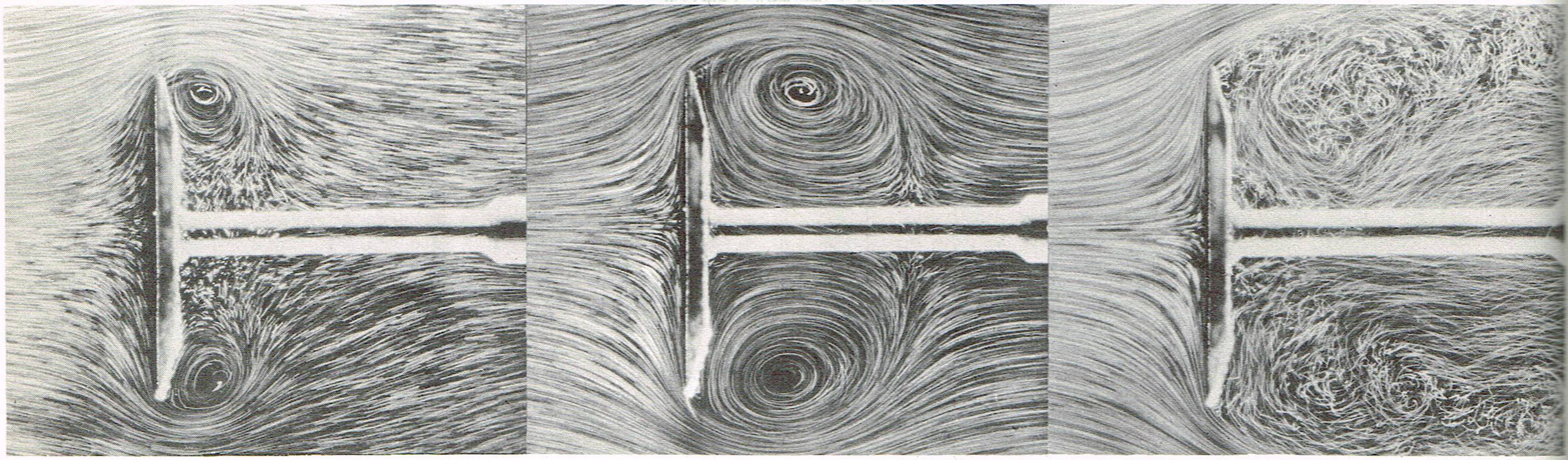
(1) Le flux sortant de S est proportionnel à la somme des charges (algébriques) intérieures.
Flux sortant (S) $= \Sigma q$.
La constante relative au vide (« permittivité du vide »), le plus naturel de tous les milieux, joue un rôle particulièrement important et, toujours épris de grec, les physiciens la désignent par ε_0 (epsilon zéro). On démontre que la constante k figurant dans la loi de Coulomb $F = k \frac{99}{2^2}$ est liée à ε_0 par la relation $4\pi k = \varepsilon_0$.



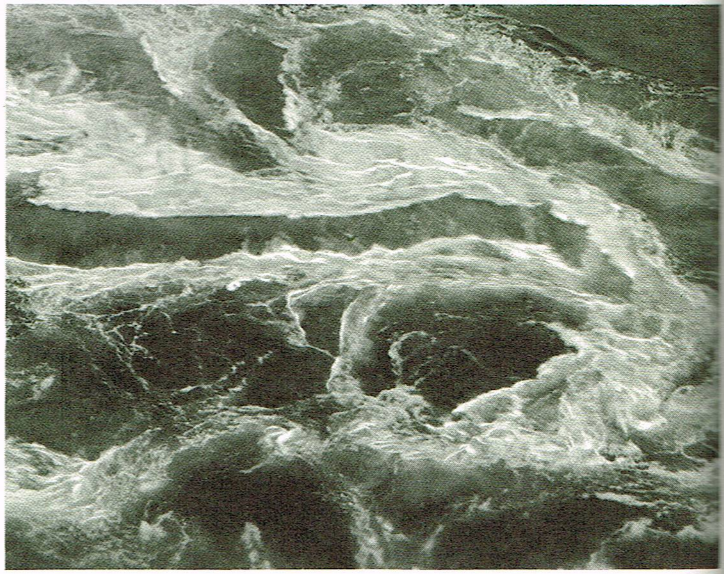
◀ Écoulement plan aux faibles vitesses avec solution colorée.

▶ Écoulement plan aux vitesses élevées avec solution colorée.





Écoulement non stationnaire. ***De gauche à droite,*** **début du démarrage, démarrage, régime permanent.**



Écoulement turbulent dans une rivière.

ne sont pas loin d'être des plans passant par l'axe du tourbillon, mais, celui-ci étant en dehors du domaine considéré, les équipotentielles sont distinctes les unes des autres et gardent sans restriction leurs propriétés. Tout change cependant si la surface envisagée contient à son intérieur une portion, si petite soit-elle, de l'axe du tourbillon : l'écoulement est radicalement différent. Le plus grave n'est pas la difficulté qu'il y a de définir correctement le flux, mais le fait qu'il ne soit plus question de trouver des surfaces équipotentielles partout distinctes, puisqu'elles viennent se couper sur l'axe du tourbillon, où elles perdent toute signification. De plus, la circulation n'est plus nulle.

Si l'on remarque que les domaines sur lesquels nous raisonnons sont aussi petits que l'on veut, il n'est pas besoin de longues réflexions pour se convaincre que l'existence du tourbillon met en jeu une propriété locale du champ et qu'on doit pouvoir attacher à ce dernier un vecteur, nul en la plupart des points, mais qui ne l'est pas sur l'axe même du tourbillon. Ce vecteur porte le nom de **rotationnel du champ.**

Il est hors de notre propos de nous étendre sur la nature et l'expression de cette grandeur. Précisons néanmoins encore quelques points. Puisqu'il s'agit d'une propriété locale, imaginons un minuscule volume de liquide brusquement solidifié, et regardons son mouvement pendant un temps très court. S'il est hors de l'axe du tourbillon, il avance : son mouvement est une translation. Mais dès qu'il est exactement sur l'axe, il pivote sur lui-même : c'est une rotation. Le « rotationnel », c'est justement cette rotation instantanée : le vecteur « rotation instantanée ».

Tournons-nous maintenant encore une fois vers le champ électrique. Chez lui, point trace de tourbillon aussi longtemps qu'il s'agit d'un champ statique indépendant du temps comme les charges électriques immobiles qui en sont les sources. Mais — que l'on nous pardonne l'expression — l'immobilité étant la mort, il faudra bien s'occuper des champs électriques « vivants », c'est-à-dire variables. Et nous devrons également envisager qu'ils aient un rotationnel différent de zéro partout ou presque partout. Peut-on se représenter un tel champ par une image hydrodynamique? Non, car c'est justement une propriété essentielle des liquides de ne pouvoir donner lieu qu'à des tourbillons localisés sur des « lignes de tourbillon », le rotationnel restant nul partout ailleurs. Il faudra donc abandonner l'espoir d'une image sensible et faire confiance à la représentation mathématique.

Phot. O. N. E. R. A. et H. Barangé.

LE CHAMP MAGNÉTIQUE

La pierre d'aimant; la boussole; le traité *De magnete* de Gilbert, qui, en 1600, inaugure la science moderne; l'étude de plus en plus précise des attractions et répulsions magnétiques, jusqu'à la loi de Coulomb qui les résume en une expression unique : voilà, esquissé à grands traits, le développement historique de la science du magnétisme.

Eh bien, ce n'est pas du tout ainsi que nous aborderons le sujet, mais, au contraire, en concentrant l'éclairage sur l'électromagnétisme, c'est-à-dire sur les actions magnétiques des courants électriques. Si étrange, en effet, que cela puisse paraître, nous ne pouvons nous défendre de considérer les aimants comme des « anomalies », bien gênantes lorsqu'il s'agit d'exposer le magnétisme.

tout courant électrique crée un champ magnétique

Le fait fondamental est donc le suivant : **les courants électriques exercent à distance des forces les uns sur les autres.** Ou, pour parler un langage plus élaboré : une charge électrique en mouvement produit dans l'espace qui l'entoure un champ capable d'action sur une autre charge en mouvement. **Un tel champ, champ magnétique, est donc lié au mouvement des charges électriques, de même que le champ électrique était lié à leur seule existence.** Plus précisément : une charge électrique immobile produit un champ électrique; si elle se meut, elle produit aussi un champ magnétique (sans pour cela cesser de produire un champ électrique).

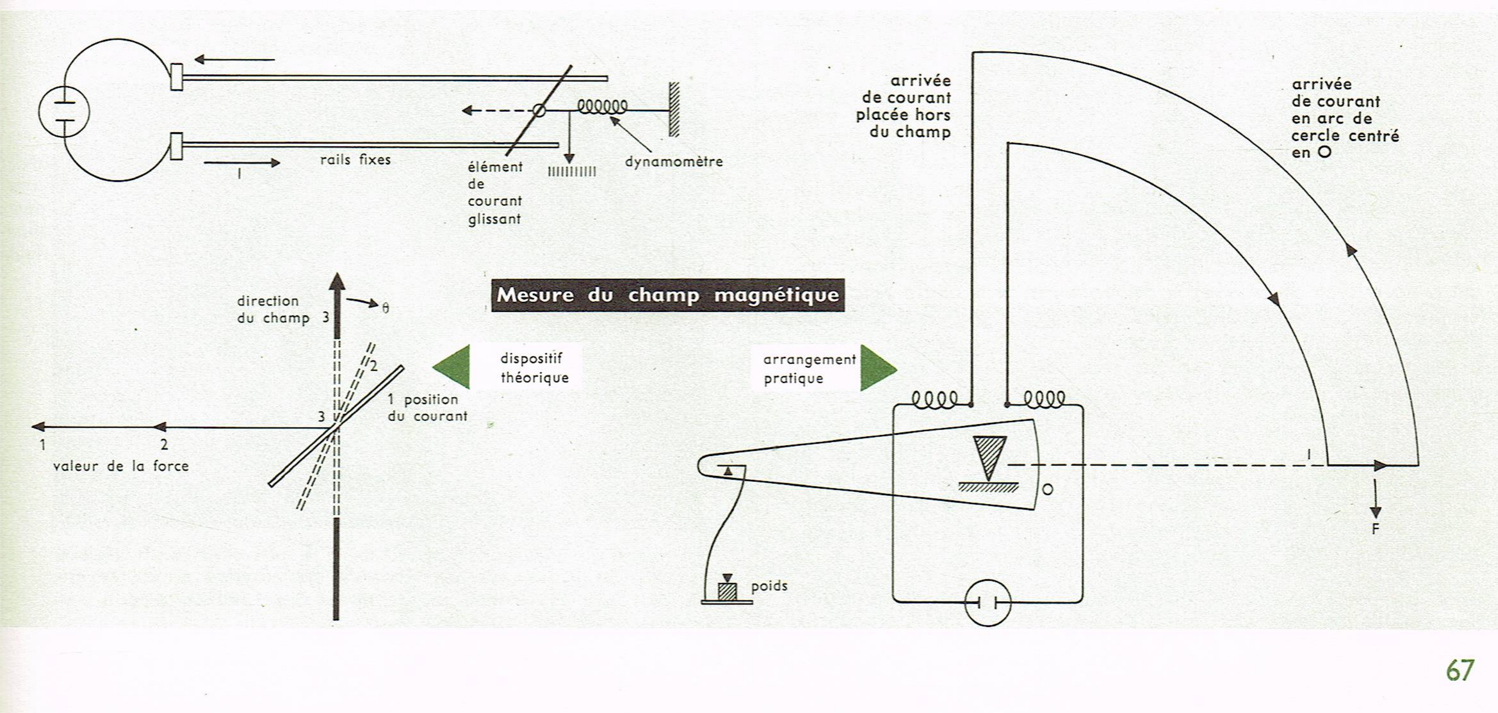
En pratique, les courants électriques permanents ne font pas apparaître de charges libres, et l'on n'observe pas de champ électrique. Ce dernier existe à coup sûr à l'intérieur du conducteur (c'est lui qui fait passer le courant malgré la résistance), mais, à l'intérieur du conducteur, nous n'y sommes pas, et dans le milieu environnant nous observons le champ magnétique seul.

Ce champ, nous ne l'avons pas encore défini avec précision, c'est-à-dire que nous n'avons pas encore donné un moyen de le mesurer, ce qui est la même chose. On mesure le champ électrique par la force qu'il exerce sur un corps d'épreuve, en l'espèce une charge électrique très petite placée au point considéré. En transposant cette façon de faire, que prendrons-nous comme « corps d'épreuve »? Il faut un courant et, afin qu'il soit localisé en un point, un élément de courant qui ne pourra naturellement pas être isolé d'un circuit, mais qui pourra fort bien être rendu seul mobile (les conducteurs d'arrivée et de sortie du courant ne l'étant pas) afin de pouvoir mesurer les forces par un dynamomètre, une balance, ou tout dispositif équivalent. « Est-ce encore une expérience mentale impraticable? » demandera-t-on. Justement, un tel circuit — astucieusement disposé de manière que les forces s'appliquant sur un élément très court soient seules mesurées — est si bien réalisable qu'il constitue l'un des dispositifs classiques de mesure du champ des électro-aimants *(balance de Cotton)*.

La force exercée en un point du champ magnétique sur un élément de courant dépend de l'orientation de ce dernier, et l'expérience seule peut nous apprendre selon quelle loi. Il y a une position du courant et une seule où la force est nulle, tandis que pour toute direction perpendiculaire à celle-ci, la force a la plus grande valeur possible. C'est la position où la force est nulle qu'il faut adopter pour la direction du champ magnétique, tout simplement parce qu'elle est unique et que c'est la convention la plus commode. Quant à la direction de la force, on trouve qu'elle est perpendiculaire à la fois au champ et au courant. Plus précisément, on trouve que la force est donnée par le produit vectoriel : $F = e\vec{v}\Lambda\vec{H}$, e étant la charge (algébrique), $\vec{v}$ la vitesse et $\vec{H}$ le champ magnétique.

quelques trajectoires d'électrons

Une catégorie de courants retiendra plus particulièrement notre attention et illustrera de façon frappante les lois que nous venons d'exprimer. Quelle catégorie? Celle des électrons en mouvement dans le vide. Raisonnons sur un seul électron : la direction du « courant » n'est autre que celle de sa vitesse. S'il est dans un champ magnétique, ce dernier exerce sur lui une force perpendiculaire, d'après ce qui précède, à sa vitesse.



Mesure du champ magnétique

Pour préciser la trajectoire, il faut regarder plus exactement la position relative de la vitesse et du champ magnétique. Si leurs directions sont perpendiculaires, un petit calcul et la simple raison de symétrie montrent que la trajectoire est plane, dans un plan perpendiculaire au champ; or, le mouvement plan où l'accélération est constamment perpendiculaire à la vitesse est le mouvement circulaire, et l'électron considéré décrit un cercle, dont le rayon R est donné par la formule

$$\left(\frac{e}{m}\right) R = \frac{v}{H}.$$

Pour un champ donné, le rayon est d'autant plus grand que la vitesse l'est elle-même. La mesure du rayon (en supposant le rapport e/m connu) donne la vitesse. De la connaissance de R, de v et de H, on peut déduire la valeur du rapport e/m. Cela, bien entendu, n'est pas limité aux électrons pris pour exemple, mais reste valable pour toute particule chargée : si deux particules de même vitesse et de charges opposées sont placées dans le même champ, elles décrivent des cercles égaux, mais en tournant dans des sens opposés. Si maintenant la vitesse n'est pas perpendiculaire au champ, la trajectoire résultante de l'électron est une hélice enroulée autour du champ magnétique, comme on peut s'en assurer en décomposant la vitesse en deux composantes, l'une perpendiculaire au champ (donnant lieu à un mouvement circulaire), l'autre selon le champ (donnant lieu à un mouvement uniforme, puisque aucune force n'y prendra naissance).

Bien loin d'être un simple divertissement géométrique, le résultat précédent a une importance tout à fait capitale en physique. **Se servir de la courbure de la trajectoire circulaire d'une particule dans un champ magnétique pour déduire sa vitesse constitue l'une des méthodes essentielles de la physique corpusculaire.** Les appareils qui rendent visible la trajectoire d'une particule isolée et qui permettent la mesure de son rayon seront décrits ailleurs. Pour le moment, bornons-nous à montrer une jolie expérience : celle que représente la figure ci-contre. Un pinceau d'électrons animés pratiquement de la même vitesse pénètre dans un ballon où subsiste assez de gaz pour que le trajet du pinceau soit rendu lumineux, et assez peu pour que la trajectoire du pinceau soit sensiblement inaltérée.

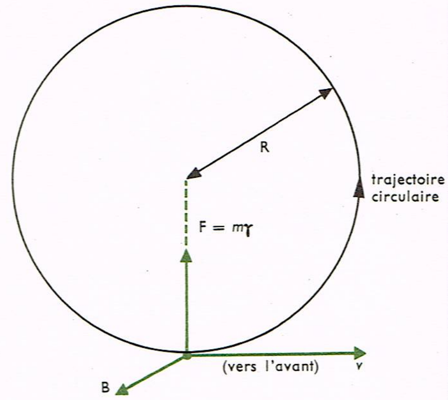
La formule donnée plus haut a encore quelque chose de très intéressant à nous apprendre : si le trajet de l'électron dans le champ magnétique n'est pas assez long pour que le cercle complet soit parcouru, un petit arc seulement de longueur l est décrit; la valeur de l'angle α entre la vitesse à l'entrée et la vitesse à la sortie du champ magnétique est donnée par la formule

$$\alpha = \frac{l}{R} = \frac{l}{v}\frac{e}{m} H.$$

La déviation varie en raison inverse de v. Or, nous avons déjà vu que la déviation d'une particule chargée par un champ électrique transversal variait en raison inverse de v^2. Par suite, si l'on fait agir en même temps sur une même particule un champ électrique et un champ magnétique sur un parcours limité et que l'on mesure les deux déviations en s'arrangeant pour qu'elles se fassent dans des directions perpendiculaires, on a un moyen de trouver le rapport e/m sans avoir besoin de connaître v. C'est particulièrement important, parce que la connaissance que l'on peut avoir de v laisse à désirer en précision et que d'ailleurs on n'a pas facilement des particules ayant toutes la même vitesse. (v. p. 70).

Des expériences de ce type ont donné la valeur que prend le rapport e/m pour l'électron; si nous nous rappelons que la charge e est connue par de nombreuses expériences, dont celles de Millikan sont un magnifique exemple (mais seulement un exemple), nous voyons comment on peut trouver la masse m, justifiant ce qui a été dit précédemment.

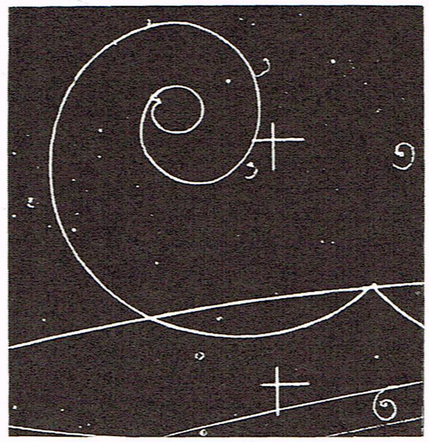
Il faut penser aussi à l'infinie variété des ions matériels positifs ou négatifs qui sont justiciables du même traitement : un champ

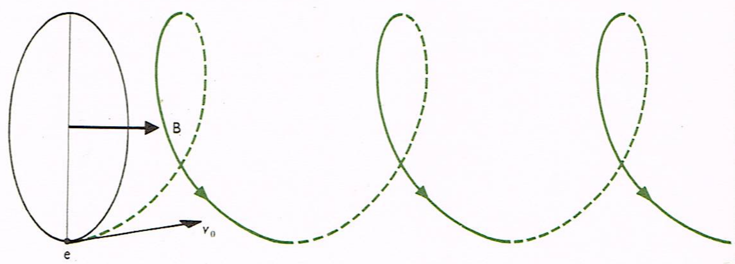


Dans le mouvement circulaire décrit par l'électron, l'accélération γ est dirigée vers le centre, donc normale au cercle et vaut v^2/R où v est la vitesse, R le rayon du cercle. $F = m\gamma$ s'écrit donc

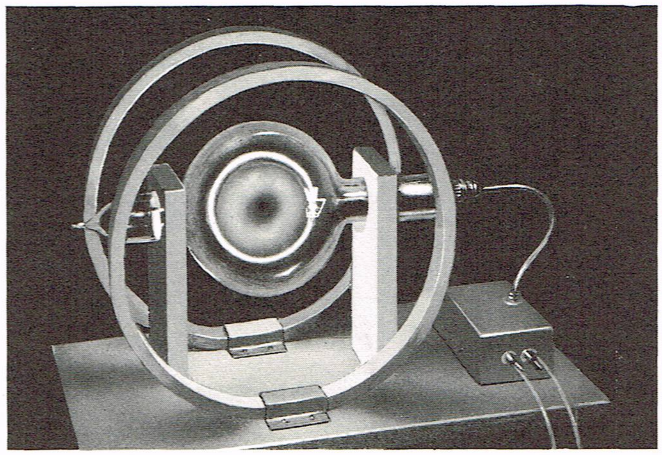
$$evH = m\frac{v^2}{R},$$

ce qui détermine le rayon du cercle.



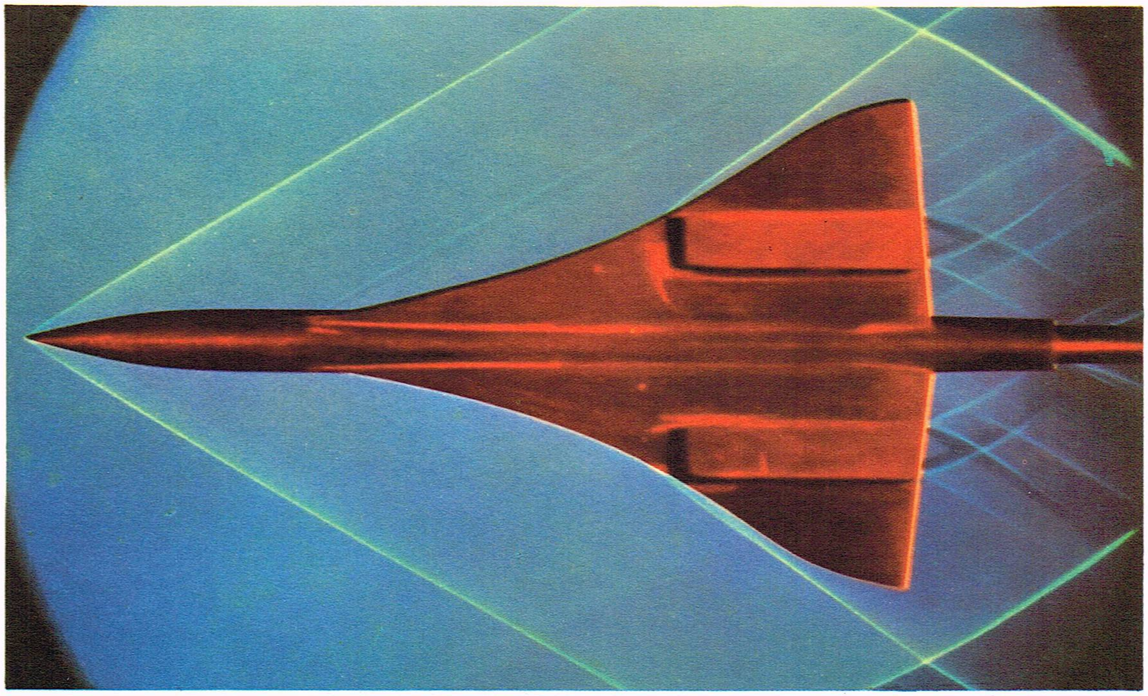


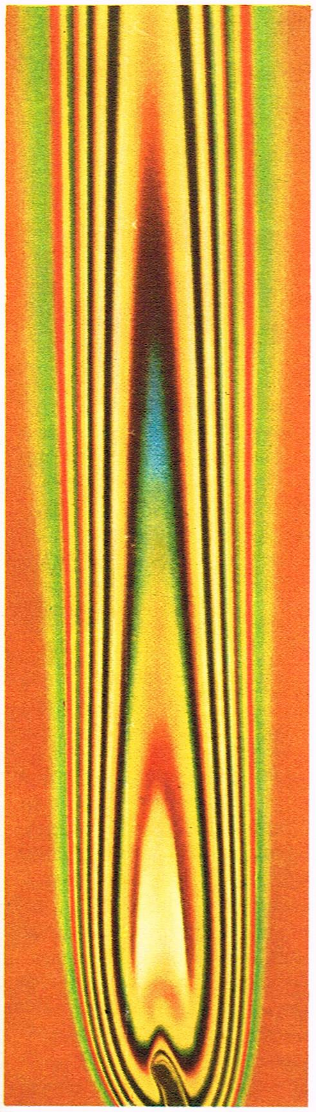
La composition d'un mouvement circulaire uniforme et d'un mouvement uniforme (selon B) se traduit par une trajectoire hélicoïdale de l'électron soumis à un champ magnétique quelconque.



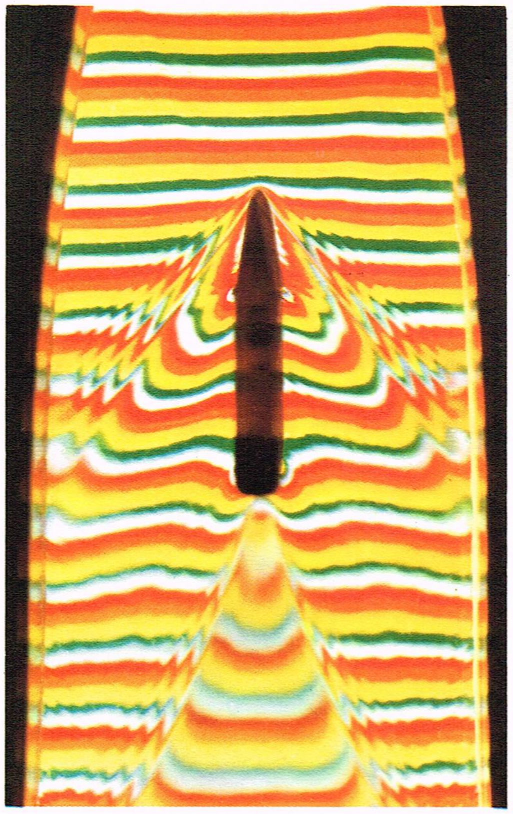
Le pinceau d'électrons produit dans le ballon et rendu visible par l'illumination du gaz résiduel est enroulé en un cercle parfait par le champ magnétique des bobines circulaires extérieures.

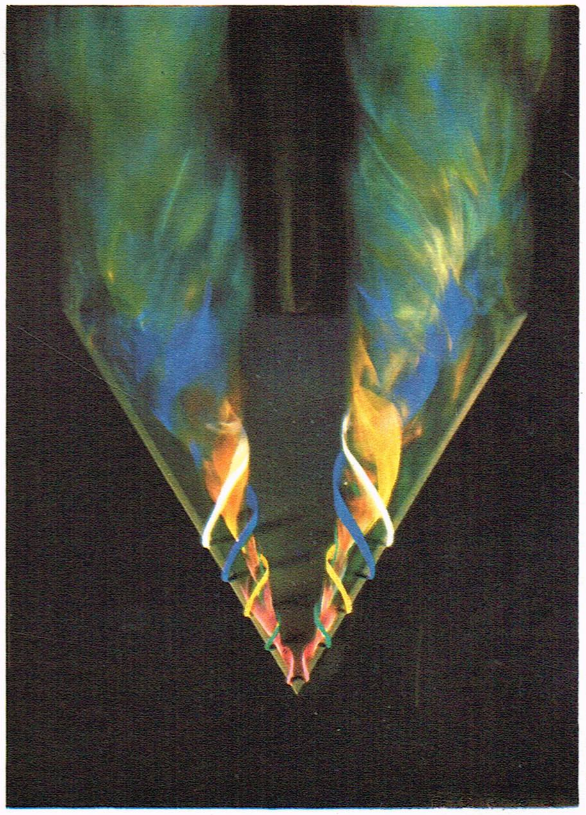
A droite, visualisation par striointerférométrie de l'écoulement autour de l'avion supersonique franco-britannique « Concorde ». *Phot.* O.N.E.R.A.





Interférométrie : flamme d'une bougie. *Phot. Zeiss.*





Ci-contre, analogie hydraulique de phénomènes aérodynamiques : écoulement à Mach 2,1; *à droite,* visualisation de l'écoulement au moyen de filets d'eau colorée. *Phot.* O.N.E.R.A.

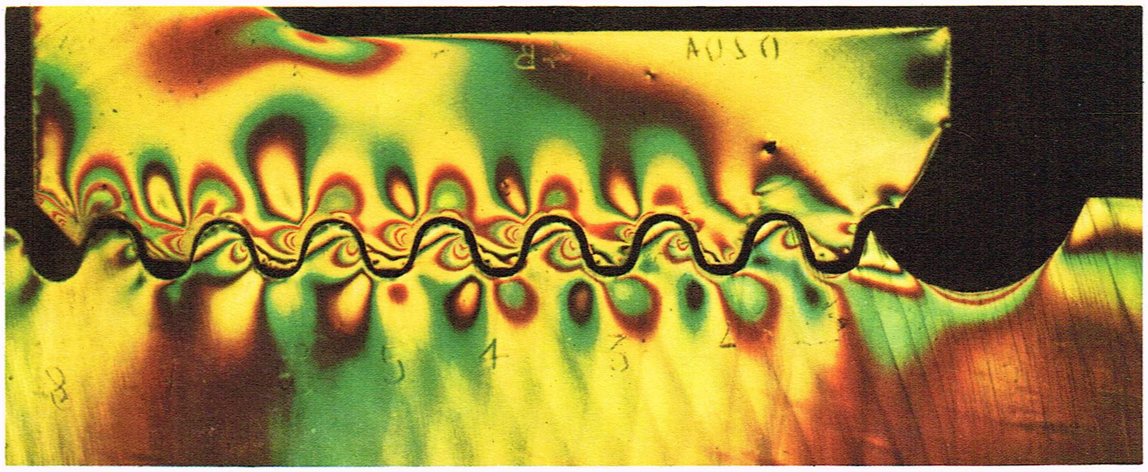
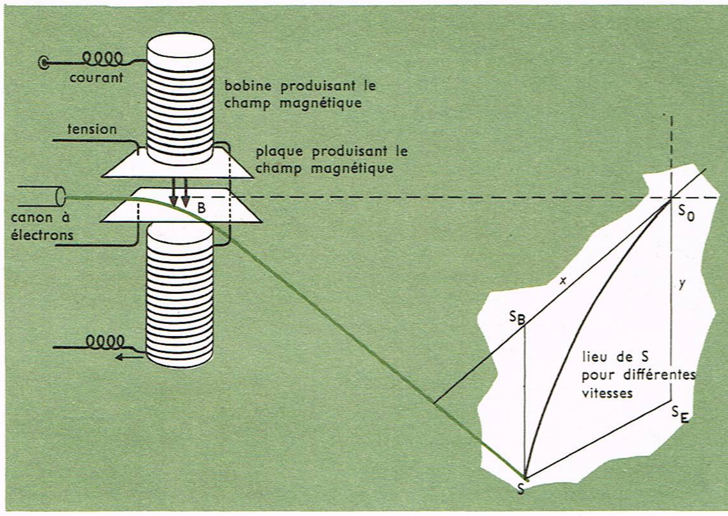


Photo-élasticimétrie : engrenages. *Phot.* O.N.E.R.A.



Déviation des électrons par un champ magnétique.

S_0 impact du pinceau d'électrons en l'absence de champ déviateur (et limite du point d'impact pour les très grandes vitesses)
S_E impact quand le champ électrique E agit seul
S_B impact quand le champ électrique B agit seul
S impact sous l'action des deux champs simultanés

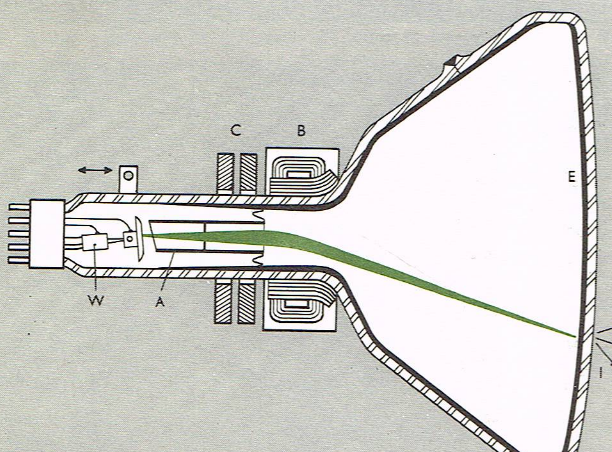
Calcul du rapport

$y = \frac{eEl^2}{m} \frac{1}{v^2}$ (déviation due au champ électrique E)

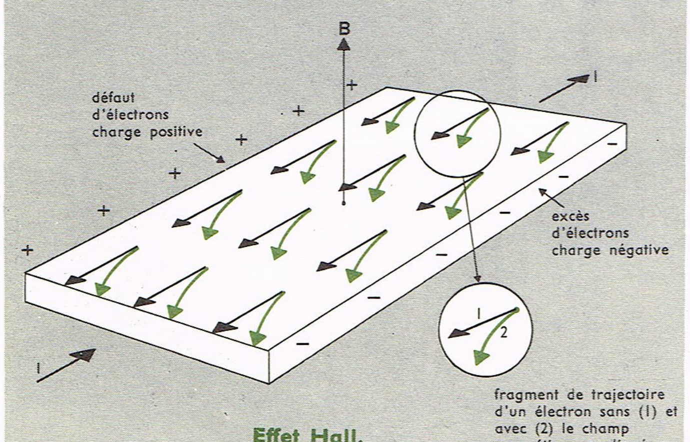
$x^2 = \frac{e}{m} \frac{B}{v}$ (déviation due au champ magnétique B).

Les deux déviations sont perpendiculaires et liées par la relation indépendante de v, $\frac{y}{x^2} = \frac{E}{B^2} \frac{m}{e}$.

Les points se rapetissent sur une parabole, que l'on observe et d'où l'on déduit la valeur de $\frac{B^2}{E} \frac{e}{m}$, donc de $\frac{e}{m}$.



Tube cathodique des récepteurs de télévision.
W cylindre de Wehnelt
A anode
C anneaux qui concentrent le faisceau d'électrons en un fin pinceau
B bobines de déviation horizontale et verticale du pinceau électronique
E écran fluorescent
I image lumineuse



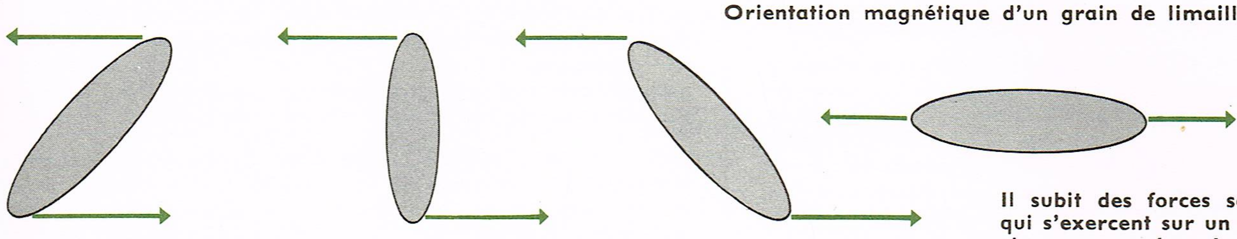
Effet Hall.

magnétique judicieusement disposé, associé ou non à des champs électriques, permet le tri des ions suivant leur charge spécifique; tous ceux qui ont le même e/m sont rassemblés, les e/m différents sont distingués. L'opération s'appelle la « spectrographie de masse »; le champ de ses applications est très vaste et embrasse tous les cas où il est utile de savoir ce qu'il y a dans un mélange gazeux.

A propos du mouvement circulaire ou hélicoïdal des électrons dans un champ magnétique, le lecteur aura suppléé de lui-même une précision importante : le champ doit être uniforme. Que dire maintenant du mouvement dans un champ magnétique plus compliqué, où la grandeur et la direction du champ sont variables? Rien de général, sinon qu'avec un champ magnétique convenable on peut tout faire d'un pinceau d'électrons, sauf augmenter leur énergie cinétique (puisque la force est toujours normale à la vitesse, et ne travaille donc pas). Ce sont notamment des champs magnétiques produits par des bobines convenablement dessinées qui assurent la concentration en un spot fin du pinceau d'électrons du tube récepteur de télévision, et la déviation de ce même pinceau qui permet de balayer l'écran. Ce sont aussi des champs magnétiques de haute intensité concentrés dans un petit volume qui réalisent les lentilles magnétiques utilisées en microscopie électronique en concurrence avec les lentilles électrostatiques. Nous avons évoqué précédemment l'optique électronique; elle fait appel tout autant à l'action du champ magnétique qu'à celle du champ électrique, et c'est par l'effet combiné de ces deux puissants outils qu'elle peut modeler à volonté les faisceaux d'électrons suivant les exigences des innombrables familles de tubes électroniques.

Donc le champ magnétique courbe en arc de cercle les trajectoires des électrons dans le vide; que fait-il aux électrons qui circulent dans les métaux? Et d'abord dans quoi est plongé l'électron mobile dans un métal sinon dans le vide qui sépare les atomes de l'édifice cristallin? On devra donc penser que là aussi les électrons voient leurs trajectoires courbées; la différence, c'est qu'au lieu d'avoir le loisir de parcourir un arc de cercle aussi étendu que le permet la région où règne le champ, ils vont bientôt heurter un atome de métal, ou au moins la surface extérieure du conducteur. S'ils sont libres de se mouvoir à l'intérieur du métal, les électrons y sont cependant emprisonnés par l'attraction collective des atomes chargés positivement. Ne pouvant donc sortir, les électrons se pressent au voisinage de l'une des faces latérales, leur concentration augmente d'un côté et baisse de l'autre; la conséquence est la naissance entre les deux surfaces d'une différence de potentiel : ce phénomène est connu sous le nom d'« effet Hall ». Même si son intensité est faible (quelques centièmes de volt), il est très intéressant, car il permet de relier assez facilement cette différence de potentiel à la densité des électrons libres

Orientation magnétique d'un grain de limaille.

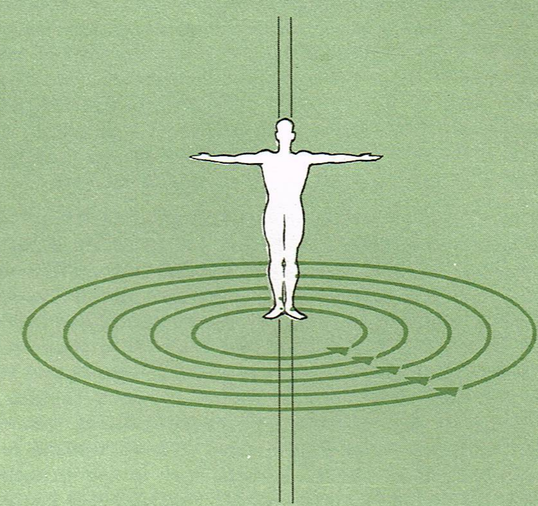


Il subit des forces semblables à celles qui s'exercent sur un « dipôle » de deux charges opposées : *à gauche*, il commence à se retourner; *à droite*, il a atteint sa position d'équilibre stable.

dans le solide conducteur, propriété qui prend toute son importance lorsque, au lieu d'un métal proprement dit, on considère un semi-conducteur comme le germanium, dont chacun sait qu'il est l'élément principal des transistors.

Il faut maintenant nous demander quelle est la structure du champ magnétique créé autour de lui par un courant circulant dans un fil conducteur, et d'abord quel est le tracé de ses lignes de force. Un moyen bien connu est de rendre celles-ci visibles au moyen de la limaille de fer. Cela revient à prendre comme « corps d'épreuve » non pas un élément de courant comme le voudrait la théorie que nous venons de développer, mais un grain ou plutôt un paillette en fer de forme allongée. Ce corps d'épreuve ne subit pas une simple force de la part du champ, mais un couple : par suite, il prend une orientation déterminée, de telle sorte que sa plus grande longueur soit disposée dans la direction du champ. Nous y reviendrons; contentons-nous pour l'instant de constater le fait : les paillettes s'orientent, s'alignent le long des lignes de force et les dessinent matériellement. Ce dessin porte le nom, consacré par la tradition, de « spectre magnétique ».

La première inspection de ces lignes de force nous montre que ce sont des courbes fermées qui entourent le fil, siège du courant; c'est un caractère essentiel, et qui indique un champ d'une tout autre nature que le champ électrostatique, avec ses lignes de force partant d'un conducteur pour aboutir à un autre sans jamais se fermer sur elles-mêmes. Il est tout indiqué de chercher une analogie dans les champs d'écoulement, et il s'en trouve tout de suite une frappante avec le champ des vitesses d'un liquide dans un tourbillon. On trouve dans l'un et l'autre des lignes de force en courbes fermées, exactement circulaires pour un tourbillon régulier d'axe vertical et pour le champ magnétique d'un courant rectiligne, presque circulaires si l'on reste au voisinage de l'axe du tourbillon ou du fil contenant le courant, plus ou moins ovoïdes dans le cas général. Quant aux propriétés précises du champ, celles qui se traduisent par des relations mathématiques, ce n'est évidemment pas à une simple inspection qu'il faut en demander l'expression et encore moins la démonstration. Celles-ci sont le fruit d'une théorie qui synthétise de nombreuses mesures. Il est permis cependant de se servir d'analogies pour les prévoir : ce champ ne comporte ni sources ni pertes, et l'on trouve qu'en effet son flux à travers une courbe fermée quelconque se conserve ou, si l'on préfère, que sa divergence est nulle partout. L'analogie avec les tourbillons fera prévoir que le champ est partout irrotationnel, sauf sur et dans les conducteurs où passe le courant. C'est justement à cet endroit que le secours de l'analogie nous abandonne. Il est, certes, tentant, puisque tous nos circuits sont constitués de fils métalliques, de raisonner comme si ces fils étaient des lignes au sens géométrique. C'est, en fait, légitime tant que l'on cherche à trouver le champ à une assez grande distance du fil (disons à une distance égale à un certain nombre de diamètres de ce dernier), mais cela ne signifie plus rien si l'on s'en approche très près, encore bien moins si l'on y pénètre. Or, il n'est pas de fil si fin qui, regardé d'assez près, ne paraisse gros comme une cheminée de navire. Chose plus grave, que dire du cas où le courant n'est pas du tout emprisonné dans un fil? Les faisceaux d'électrons circulant dans les tubes à vide sont dans ce cas, et il est des exemples plus grandioses imposés



Le sens des lignes de force est relié à celui du courant par la « règle du tire-bouchon ». On l'enfonce dans le sens du courant en le tournant dans celui du champ. On peut préférer la « règle du bonhomme d'Ampère »; il voit les lignes de force tournées vers la gauche.

par la nature : si nous levons les yeux vers le Soleil, il nous fera songer aux énormes flux de particules électrisées qu'il émet, sans parler des courants qui brassent sa propre matière.

Le problème du champ magnétique au sein des courants ne peut donc pas être éludé; il n'est pas accessible à l'expérimentation directe et demande le secours d'une théorie mathématique qui n'a pas sa place ici. Donnons-en seulement quelques résultats : **le champ a en chaque point un rotationnel qui est proportionnel au vecteur densité de courant** (le rapport entre le rotationnel du champ et la densité de courant est une constante particulière à chaque milieu appelée « perméabilité »). Une importance spéciale s'attache à la valeur qui correspond au vide, ou — pratiquement — à l'air qui est désignée par μ_0. Cette valeur numérique dépend naturellement des unités choisies pour I et H; dans le système M. K. S. A. elle vaut $4\pi.10^{-7}$. **La divergence reste partout nulle, ainsi que la circulation.**

Ces résultat paraîtront sans doute décevants en ce qu'ils ne permettent pas de « représentation » tombant sous les sens; ils seront essentiels à l'intelligence des rapports entre champs magnétique et électrique.

et si l'on parlait un peu des aimants ?

« Ces relations constituent l'essentiel de l'électromagnétisme », conclut le théoricien. « Vraiment? », pourrait lui répondre quiconque a les yeux ouverts sur les choses. « L'électromagnétisme c'est au moins autant les propriétés du fer et de ses alliages, qui rendent possibles toutes les machines. Pensons aux transformateurs, aux machines génératrices, aux moteurs, à l'innombrable famille des électro-aimants. Les plus petits animent des relais gros comme l'ongle, tandis que les géants de l'espèce imposent leur masse énorme aux synchrocyclotrons et aux grands accélérateurs de particules. »

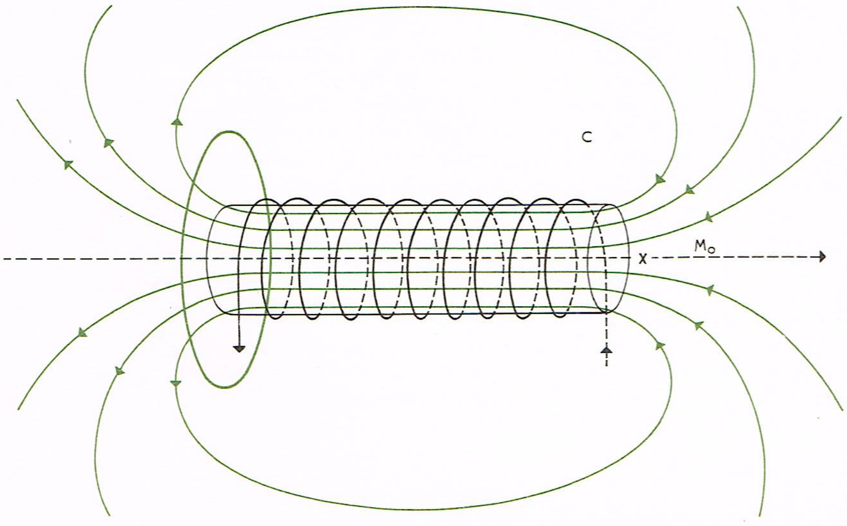
L'électro-aimant, c'est le fer associé au courant électrique, mais que penser des aimants permanents? Tantôt serviteur modeste qui ferme une porte d'armoire ou retient le cendrier sur la table malgré le roulis du bateau, tantôt promu à une fonction essentielle dans le compas magnétique de navigation, l'aimant joue des rôles infiniment variés dans la technique et propose obstinément à la physique l'énigme de son existence. Premier point à bien préciser : le champ magnétique est toujours identique à lui-même quelles que soient ses origines. Très exactement, si l'on constate qu'une boîte hermétiquement close est entourée d'un champ magnétique, il n'est aucun moyen de savoir si cette boîte contient un aimant ou une bobine parcourue par un courant (avec ou sans noyau de fer).

En définitive, ce sont les propriétés étonnantes du fer qui constituent le nœud de la question. Un fil enroulé en hélice et parcouru par un courant — une bobine — produit un certain champ magnétique. Faisons le même enroulement sur un noyau de fer : avec le même courant, le champ magnétique observé en M (voir la figure) au voisinage d'une face est facilement 1 000 ou 2 000 fois plus grand. Le flux qui passe à travers une section voisine C est augmenté dans le même rapport. Quand nous aurons dit que le fer possède une **perméabilité** très élevée à l'égard du champ magnétique, nous aurons nommé le phénomène. Quand nous aurons précisé que le fer concentre à son intérieur les lignes de force magnétique, les attire en quelque sorte sans modifier essentiellement leur caractère de courbes fermées, nous aurons ajouté un utile commentaire au mot de « perméabilité »; mais nous n'aurons toujours rien expliqué.

Si notre boîte contient un aimant, nous constatons tout de suite que les lignes de force ne sont pas fermées, qu'elles vont d'un pôle à l'autre. Si plusieurs aimants sont en présence, elles vont encore d'un pôle à un pôle, qu'ils appartiennent au même aimant ou à deux aimants différents. Les pôles « nord » fonctionnent donc comme des sources du champ, et les pôles « sud » comme des pertes, comme s'il y avait des charges (ou masses) magnétiques positives et négatives.
Tout cela n'est-il pas en flagrante contradiction avec ce que nous avons vu des champs magnétiques, sans perte ni source? A cela, que répond l'électricien? Qu'il ne faut pas se fier aux apparences, qui, en l'espèce, sont fallacieuses; que les « pôles » ne correspondent à rien de réel; qu'il n'y a pas de masses magnétiques; que les lignes de force sont bel et bien fermées sur elles-mêmes, fermées à l'intérieur de l'aimant. Halte-là!

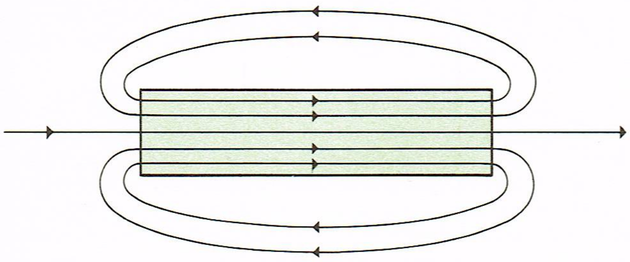
Une chose est d'affirmer, autre chose de prouver. Le champ électrique d'un bâtonnet de verre ou de soufre électrisé et le champ magnétique d'un aimant cylindrique sont absolument semblables; pourquoi donc les charges seraient-elles là réalité, ici apparences qualifiées de fallacieuses? Et si l'on dénie toute existence aux charges magnétiques, que propose-t-on pour les remplacer, car il ne peut exister de courant, au sens ordinaire, dans un aimant? Un tel courant, en effet, ne peut subsister sans une dépense d'énergie, sans qu'une pile ou quelque chose d'équivalent s'use ou se décharge. Rien de tel dans un aimant, qui reste aimanté indéfiniment.

La réponse à la question relative aux charges est assez facile. L'électrostatique nous a montré comment, à partir des corps électrisés — peu importe comment —, on arrive à isoler des parties portant une charge d'un signe ou de l'autre, dont la goutte d'huile des expériences de Millikan est le type. En étudiant ces corps chargés, on décèle par la mesure et ensuite on isole effectivement l'électron, la charge élémentaire. Le même effort d'analyse appliqué aux corps aimantés n'aboutit pas du tout au même résultat : **quoi que l'on fasse, on ne trouve jamais que des aimants ayant deux pôles.** Les minuscules paillettes qui dessinent les spectres magnétiques sont précisément de cette sorte. Comment s'en aperçoit-on? Parce que, dans un champ uniforme (et si la paillette est assez petite, tout champ — sauf cas exceptionnel — peut être considéré comme approximativement uniforme à son échelle), elles subissent une orientation et non un déplacement, en d'autres termes sont soumises à un couple et non à une force. Tous les efforts pour isoler un « pôle magnétique » d'un type déterminé sont restés vains. Ces efforts furent nombreux et ingénieux; entre 1944 et 1948, pour ne parler que des plus récents, des expériences fort curieuses furent faites par le physicien américain Ehrenhaft : dans certains aérosols soumis à un éclairement intense et placés dans un champ magnétique, on voyait des particules solides paraissant se comporter comme des pôles magnétiques. Ce résultat était troublant, mais les expériences refaites avec le plus grand soin par Tauzin et quelques autres physiciens livrèrent leur secret, et il fut établi que les mystérieux objets n'étaient rien d'autre que des grains métalliques dont tous n'étaient d'ailleurs pas magnétiques. Lorsqu'ils sont magnétiques, ce sont de petits aimants à deux pôles, tout comme les paillettes qui servent à produire les spectres magnétiques, à la grosseur près. C'est le choc des molécules de l'air combiné avec l'échauffement produit par l'éclairage intense qui a été montré responsable de leurs mouvements paradoxaux.

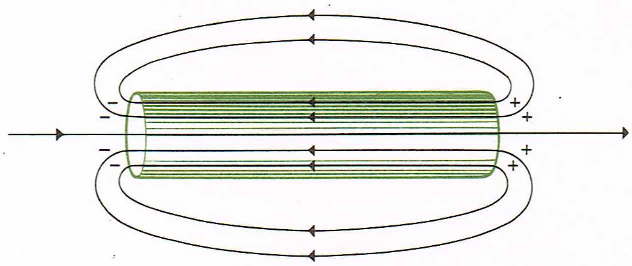


Bobine avec noyau de fer.

Le champ en M_0, le flux à travers C sont, par rapport à leurs valeurs en l'absence de noyau, multipliés par un facteur peu inférieur à μ. Plus précisément : le champ serait multiplié par la perméabilité (ou plus exactement par le rapport μ/μ_0) si le circuit était plongé dans un milieu *indéfini* ayant cette perméabilité. Le calcul précis du champ en présence du noyau de fer est assez compliqué : c'est un important problème d'électrotechnique. Dans le cas de notre figure, le champ est à grande distance peu modifié par l'introduction du noyau, mais en un point tel que M_0, il est pratiquement multiplié par μ/μ_0.



Deux champs en apparence semblables.

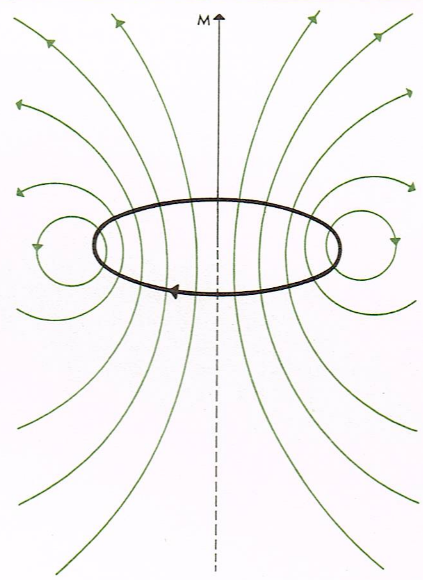


Découvrira-t-on un jour les fameux pôles isolés? Impossible n'est pas un mot de la physique, mais il ne nous paraît pas hasardeux de donner cette éventualité comme hautement improbable.

Ce n'est donc pas par décision arbitraire, mais uniquement par soumission aux faits et aux résultats expérimentaux que les physiciens accordent la réalité aux charges électriques et la refusent aux charges magnétiques. Et si l'on regarde de plus près des champs en apparence « absolument semblables » (voir figure ci-dessus), c'est pour y découvrir des différences irréductibles dans le comportement que l'on doit attribuer au champ à l'intérieur des corps solides. A l'extérieur, comme au sein du bâtonnet isolant électrisé, le champ va des charges positives aux charges négatives; tandis qu'au sein du corps aimanté, il conserve la même direction en continuité avec l'extérieur, de manière à obéir à la conservation du flux sur toute la longueur des tubes de force.

Ainsi, c'est entendu : pas de masses magnétiques. Mais comment alors expliquer à la fois les propriétés et les actions des aimants permanents, et la polarisation du fer « doux » : car nous sentons bien qu'il faut une explication unique? Une solution satisfaisante et, semble-t-il, définitive, a été vue dans le deuxième quart de ce siècle, mais pour comprendre de quoi il s'agit, il faut remonter assez loin dans le passé. Ampère, dont la pensée était guidée par la vision d'une unité profonde du magnétisme des courants et des aimants, admettait l'existence dans ces derniers et dans le fer de « courants particulaires » responsables des effets observés, et indiquait une méthode pour les calculer. Imaginons, en effet, un petit courant circulaire (sans nous demander comment, en l'absence de pile ou de tout autre générateur, ce courant peut persister indéfiniment), l'ensemble de ses lignes de force est celui d'un petit aimant; placé dans un champ uniforme, il s'oriente perpendiculairement au champ. En imaginant un nombre suffisant de telles boucles, ou « dipôles magnétiques », on peut, en principe, « reconstituer » les effets extérieurs de n'importe quel corps aimanté, aussi bien temporaire que permanent. Pour obtenir effectivement la répartition de courants particulaires correspondant à une aimantation observable, il faut une hypothèse sur la structure intime du milieu aimanté. A l'époque d'Ampère, ou la réalité atomique était loin d'être admise par tout le monde, on n'avait pas les bases nécessaires pour trouver une structure plausible, et les « courants particulaires » gardèrent longtemps le caractère d'une fiction analytique.

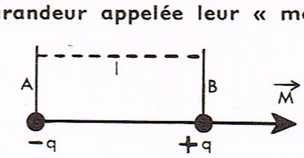
Nous sommes autrement favorisés aujourd'hui. Nous connaissons l'arrangement des atomes formant les cristaux et leurs structures intérieures; pouvons-nous donc voir les courants particulaires? La première réaction est de répondre : oui, naturellement; les courants ne sont pas autre chose que les électrons tournant autour des noyaux sur leurs orbites planétaires... Seulement, cette idée, séduisante au premier abord, se révèle insoutenable quand on essaie d'en développer les conséquences.



Boucle de courant ou dipôle magnétique.

La ressemblance des lignes de force avec celles d'un aimant est d'autant meilleure qu'on est plus loin de la boucle. Sur la figure ci-dessus, on reste beaucoup trop près.

Cette expression, copiée du dipôle électrique qui, lui, est bien formé de deux charges opposées séparées par une distance *l*, ne doit pas nous tromper. Elle est bien commode parce qu'elle nous rappelle que ces deux objets se traitent de la même manière et se caractérisent par une grandeur appelée leur « moment ».



Ce dernier vaut, pour le dipôle électrique, le produit ql et il est dirigé suivant AB; pour le dipôle magnétique, il vaut le produit IS et est dirigé suivant la normale au plan de la boucle. La règle d'équilibre veut que dans tous les cas l'angle du moment et du champ soit le plus petit possible, c'est-à-dire nul si le dipôle est capable de tourner librement.

Il est parfaitement exact qu'un électron décrivant une orbite est un petit courant et, par conséquent, un petit aimant. Pour les atomes ayant de nombreux électrons, les différentes trajectoires s'arrangent quant à la forme, à la position et au sens du parcours de telle sorte que l'aimantation résultante soit minimale et si possible nulle. Fort bien, dirons-nous, le magnétisme sera donc l'apanage des atomes qui possèdent une aimantation importante résultant de la disposition exceptionnelle de leurs orbites électroniques. Parmi ces derniers, pourront donner des aimants permanents ceux qui, par un mécanisme à élucider, seront capables de conserver leur aimantation individuelle parallèle dans un volume fini de matière.

Mais on ne peut pas avoir un modèle atomique destiné à expliquer les propriétés magnétiques, un autre qui aurait pour fonction de rendre compte de la conductibilité, un autre encore pour d'autres propriétés. Or, les modèles issus de la spectroscopie, particulièrement convaincants par la perfection avec laquelle ils expliquent l'absorption et l'émission de la lumière, excluent précisément cette aimantation liée à des orbites électroniques. Au contraire, les propriétés spectroscopiques et le magnétisme s'unissent pour imposer l'idée que c'est **l'électron lui-même qui est un minuscule aimant.** Indépendamment donc de sa charge électrique, il possède une aimantation propre, associée d'ailleurs à une quantité de mouvement de rotation propre, comme si l'électron était une sorte de toupie ou de fuseau en rotation; cette propriété, désignée sous le nom de *spin* (mot anglais signifiant « toupie » devenu d'un emploi universel), est un des caractères les plus remarquables et les plus mystérieux de l'électron, qui se trouve être bien loin de la simplicité de la « charge ponctuelle » évoquée plus haut. Cette propriété sera étudiée d'une manière approfondie dans les chapitres consacrés à la physique corpusculaire.

les trois formes de magnétisme

Nous pouvons donc nous représenter ainsi, dans un résumé excessivement sommaire et simplifié, les propriétés magnétiques de la matière; molécules (exceptionnellement atomes libres) dans les gaz et les liquides, ions ou molécules dans les cristaux peuvent avoir trois comportements principaux. Dans la plupart des cas, il s'établit une compensation exacte du magnétisme propre des électrons, de sorte qu'à l'état naturel la molécule est privée de tout magnétisme. Un champ magnétique appliqué a néanmoins une petite action, car il déforme légèrement les orbites électroniques, causant une très faible aimantation. On peut prévoir et vérifier qu'elle est en sens inverse du champ magnétisant; c'est le **diamagnétisme.** Pour fixer un ordre de grandeur, si l'on prend 1 pour la perméabilité du vide, celle d'un corps diamagnétique en diffère de l'ordre de un cent-millième en général.

Les corps **paramagnétiques** sont constitués de molécules possédant chacune une aimantation propre (généralement celle d'un électron non compensé), mais qui sont orientées dans le plus parfait désordre. Un grain, même très petit, contenant un nombre énorme de molécules, l'aimantation résultante est, en moyenne, nulle. Vienne un champ magnétique; il tend à orienter dans sa direction les aimants moléculaires, que les chocs et les rebondissements, incessants au sein de la matière, tendent à remettre en désordre. Un compromis s'établit, qui laisse une légère aimantation dans le sens du champ et, par suite, une perméabilité qui dépasse un de quelques millièmes ou un peu moins. On sait que le froid restreint l'agitation moléculaire, donc favorise l'orientation par le champ. L'aimantation paramagnétique devient intense aux basses températures, c'est-à-dire à celles que réalisent l'hydrogène ou l'hélium liquides.

Le cas des **ferromagnétiques** (fer et nickel en sont les exemples les plus évidents, mais il y en a plus qu'on ne croit) est le plus étrange. Non seulement les spins ne se compensent pas dans chacun des ions, mais il possèdent une sorte de discipline spontanée qui leur fait assumer une orientation commune dans un volume, petit, mais contenant un grand nombre d'ions. Ces « domaines » sont petits à notre échelle : on peut (ce n'est pas facile) réduire du fer en une poudre assez fine pour que chaque grain soit un domaine. En l'absence de champ directeur, les domaines sont orientés au hasard, et aucune aimantation ne se manifeste pour un morceau contenant un grand nombre de domaines. Mais il suffit d'un champ très modéré pour les orienter pratiquement tous, faisant apparaître une aimantation considérable. La perméabilité atteint ici plusieurs milliers avec le fer, plus de 100 000 dans les alliages « hautement perméables ».

Dans les aciers et divers alliages complexes dont les domaines sont en majorité orientés, interviennent des effets (encore mystérieux) ayant quelque analogie avec le frottement; l'orientation survit alors au champ qui l'a causée, et l'on a des aimants permanents.

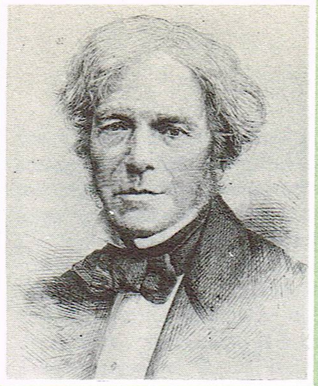
Le fait que le magnétisme, un des phénomènes les plus anciennement observés, ne trouve son explication qu'à la lumière des propriétés de l'électron les plus difficiles à comprendre est un sujet d'intéressantes méditations.

L'INDUCTION

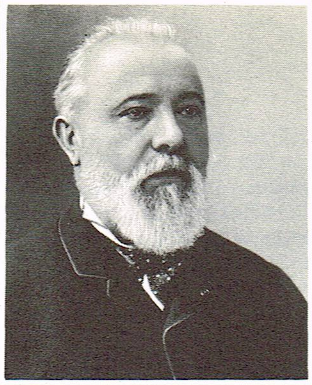
Au contraire des chapitres précédents, l'éclairage historique paraît ici particulièrement suggestif. On connaissait déjà au XVIII[e] siècle l'électrisation sans contact, consistant dans l'apparition simultanée des charges des deux espèces sur un corps placé au voisinage d'un pôle d'une machine électrique. En langage moderne, nous disons que le corps a été *polarisé* par le champ électrique; on disait alors qu'il avait subi l'influence (ou, en anglais, l'*induction*) de la source d'« électricité de tension ». Dès qu'on eut découvert l'« électricité en mouvement » (le courant électrique), on s'est demandé si elle n'était pas capable, elle aussi, d'induire des effets, plus précisément d'induire de l'électricité de même « nature », c'est-à-dire un courant. Rapidement on se rendit compte que cet effet, s'il existait, n'était pas très facile à voir; les expérimentateurs multiplièrent les dispositifs, les arrangements ingénieux du circuit inducteur et des conducteurs destinés à subir l'induction, et personne ne vit rien.

Parmi ces chercheurs, il y avait Faraday, et Faraday non plus ne voyait rien : « Extrait de mon cahier de notes en date du 28 novembre 1825 : expérience sur l'induction par le fil réunissant les bornes d'une batterie voltaïque : pôles réunis par un fil d'environ 4 pieds, parallèlement auquel était monté un autre fil séparé du premier par une double épaisseur de papier, et dont les extrémités étaient attachées à un galvanomètre. Aucune action manifestée; impossible de mettre en évidence une induction de la part du fil parcouru par le courant. »

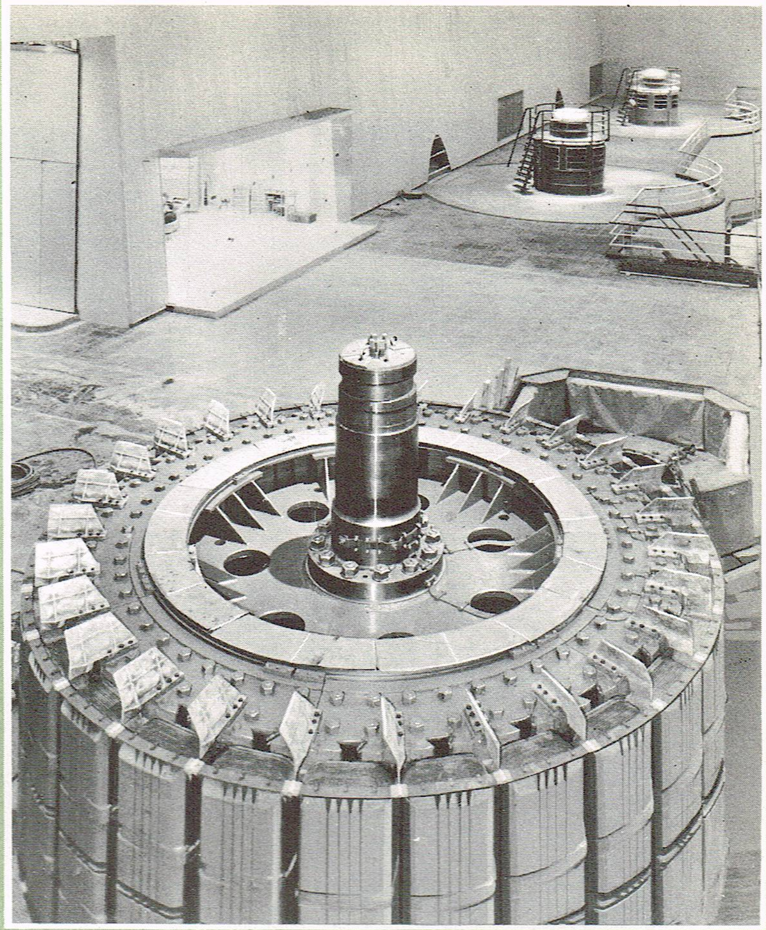
Il continua cependant à chercher et, en 1831, il vit ce qu'il fallait voir. Contrairement à beaucoup d'ouvrages scientifiques anciens, les écrits de Faraday sont non seulement lisibles, mais véritablement passionnants : on y voit la science se faire de page en page, dans un style clair et alerte, sans aucun recours aux mathématiques (ce qui n'est pas toujours pour simplifier). Aussi ne résisterons-nous pas au plaisir de le citer de nouveau :



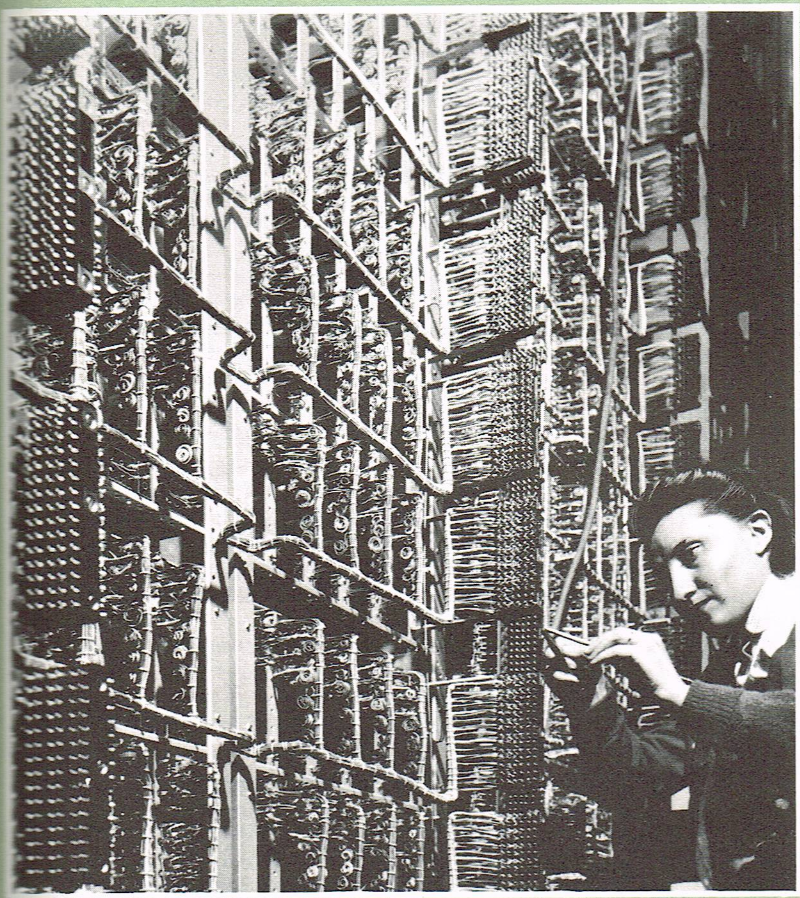
Michael Faraday
(1791-1867)



Zénobe Gramme
(1826-1901)



Alternateur à axe vertical de 90 000 kVA

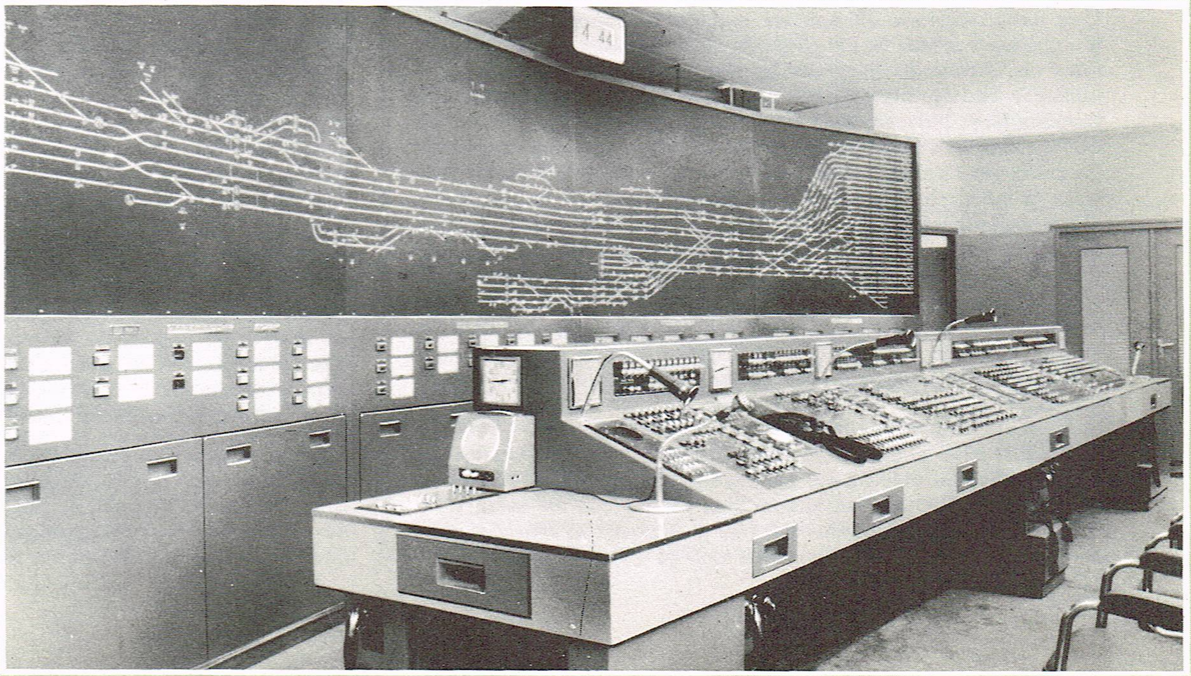


Détail d'un central téléphonique.

La loi de Faraday et quelques-unes de ses applications.

$$V = -\frac{d\Phi}{dt}$$

Le signe — rappelle que les effets d'induction tendent toujours à s'opposer aux variations qui les produisent. Si le flux est exprimé en unités MKSA (« weber »), cette formule donne la tension V en volts.



Poste d'aiguillage, dit « poste à billes », commandant un faisceau de triage, à manœuvre électrique, dans lequel les wagons sont représentés par des billes qui établissent, sous le contrôle de relais d'occupation de zones, les contacts nécessaires à la manœuvre des aiguilles.

Phot. Bibl. nat., Petit, Electricité de France, Belzeaux. S. N. C. F.

« 203 pieds de fil de cuivre ont été enroulés sur un cylindre de bois; 203 pieds de fil semblable étaient intercalés entre les spires de la première bobine, le contact électrique entre les deux fils étant empêché par de la ficelle. Une des hélices était connectée avec un galvanomètre, l'autre avec une batterie... Quand le contact était établi, il y avait un **rapide et très léger effet** sur le galvanomètre, et un léger effet semblable quand le circuit était interrompu. Mais aussi longtemps que le courant voltaïque continuait à passer dans une bobine, on ne pouvait avoir aucune action galvanométrique ni aucun effet tel qu'une induction sur l'autre hélice... »

En sachant apercevoir cet « effet rapide et très léger », Faraday venait de réaliser une découverte capitale entre toutes : la condition nécessaire pour qu'un courant produise des effets d'induction est qu'il soit **variable.** Il a été reconnu ensuite que l'effet était d'autant plus grand que le courant était plus rapidement variable ou — pour parler le langage mathématique, seul assez précis — que l'effet était proportionnel à la dérivée du courant par rapport au temps.

Notons en passant que l'arrangement originel de Faraday était une ébauche de transformateur, auquel manquait seul le noyau de fer. Celui-ci d'ailleurs fut ajouté par Faraday lui-même dans une expérience ultérieure qui lui montra des effets d'induction beaucoup plus puissants. En même temps, il découvrit que les mêmes effets se produisaient lorsqu'on approchait ou qu'on éloignait un aimant d'un circuit; ou encore quand on déplaçait, par exemple en le faisant tourner, un circuit dans le champ d'un aimant fixe. Dès lors, bien que le mot « induction » ait été pieusement conservé, on s'est trouvé bien loin de ce que l'on cherchait au départ : une induction d'un courant par un courant. La cause inductrice n'est pas le courant, mais le champ magnétique, qu'il provienne d'un courant ou d'un aimant; ou, plus exactement, ce n'est pas le champ, mais ses variations ou ses déplacements. Qu'il s'agisse d'un champ mobile et d'un circuit fixe, ou d'un circuit mobile dans un champ constant, ou des deux mobiles ensemble, l'analyse de tous les cas montre qu'on peut toujours considérer une seule et même grandeur dont la variation est responsable de l'effet d'induction, et qui est le flux du champ magnétique à travers le circuit.

Il faut encore ajouter une précision avant d'arriver à la loi générale. Nous avons parlé de courants induits; qu'arrive-t-il si le circuit soumis au champ variable n'est pas tout à fait fermé? Aucun courant ne peut passer, mais entre les deux extrémités en regard, il se développe une tension (si elle est assez élevée, elle produit une étincelle; et alors le circuit est fermé par cette étincelle, car celle-ci est un conducteur).

la loi de l'induction

Maintenant, l'énoncé suivant doit sembler limpide : **toute variation de flux magnétique à travers un circuit y fait apparaître une tension (ou force électromotrice), qui est proportionnelle à la dérivée du flux par rapport au temps.** C'est par l'intermédiaire des phénomènes d'induction que l'électricité domine le plus notre vie quotidienne : c'est, en effet, toujours par des machines à induction que l'énergie électrique industrielle est produite. Ce sont des transformateurs qui en permettent la distribution. Ce sont des moteurs à induction qui animent toutes les machines-outils. Ce sont des bobines d'induction qui allument les moteurs à explosion. Ce sont des effets d'induction qui jouent un premier rôle dans tous les systèmes de télécommunications. Assez curieusement, toutes nos machines tournantes, des plus minuscules aux plus géantes, sont la postérité de la dynamo Gramme. Zénobe Gramme, l'inventeur, était l'un de ces génies intuitifs sans culture scientifique et qui ne savait certainement pas ce qu'est une dérivée. Comprenons bien, cependant, que sa machine primitive n'a eu une telle fécondité que grâce aux soins de générations d'ingénieurs et d'hommes de science doués, eux, de culture mathématique.

On a pu faire des romans de science-fiction, on peut en faire encore sur le thème suivant : un jour le fer est brusquement privé de ses propriétés magnétiques. (Disons, par exemple, que les spins décident une grève illimitée.) Aussitôt, les effets d'induction sont considérablement réduits, alternateurs et transformateurs deviennent impuissants. Plus de moteurs électriques; plus de moteurs à explosion non plus, définitivement dépourvus d'allumage. Les téléphones sont muets, l'électronique paralysée. La civilisation s'écroule dans des convulsions. Les survivants retournent à l'âge des cavernes, jusqu'à ce qu'un chasseur, en frottant l'ivoire d'une dent sur une peau de bête, réinvente une machine électrostatique et qu'un nouveau cycle scientifico-industriel reparte sur cette base.

LES ÉQUATIONS DE MAXWELL

Que nous utilisions ou non les machines de notre civilisation, nous nous voyons, pour peu que nous réfléchissions, plongés dans un ensemble complexe de champs électriques et magnétiques : champ magnétique terrestre, dont l'exploration de l'espace nous permet de soupçonner à quelle distance il s'étend autour de la Terre; champs électriques dans l'atmosphère; potentiel des nuages orageux se montant à des centaines de millions de volts; champs électriques associés aux flux de particules électrisées projetées par les étoiles, sur toute la Galaxie; champs magnétiques produits par les mouvements des nuages que ces mêmes particules constituent dans l'espace et dont les volumes sont plus grands que le système solaire. Les uns et les autres agissant à la manière de gigantesques accélérateurs, dont la chambre à vide n'est autre que l'espace illimité, se manifestent à nous par des particules de rayonnement cosmique dont l'énergie se chiffre par de nombreux milliards d'électron-volts. Plus audacieusement encore, essayons de pénétrer dans le très petit, plus précisément dans l'espace vide qui occupe la quasi-totalité du volume des solides les plus compacts, dès qu'on les aborde à l'échelle atomique : cet espace est, lui aussi, parcouru en tous sens par des champs électriques et magnétiques. Si, anticipant enfin sur ce qui sera exposé plus loin, nous tenons compte de ce que rayonnement et lumière sont aussi des champs de cette nature, alors ceux-ci apparaîtront bien comme des constituants essentiels de l'univers.
Deux constituants distincts ou un seul constituant? **Qu'il y ait des relations intimes, universelles, entre champ électrique et champ magnétique, pas besoin d'être grand prophète pour le pressentir; autre chose est de le formuler.**

Chercher quelles sont ces relations sur le plan de l'être est une entreprise du genre de celles qui tentèrent les philosophes — depuis Pythagore jusqu'à Kant ou Hegel — en quête du « savoir absolu ». Recherche que la science moderne (non sans regret — mais c'est une autre histoire) a abandonnée, se contentant des relations sur le plan opérationnel, c'est-à-dire celui qui extrait l'enseignement des expériences passées pour en faire le guide des expériences futures.

C'est assez dire que ces relations s'expriment bien par la mathématique et mal par le discours, c'est pourquoi on les nomme « équations de Maxwell ». Juste hommage rendu au fondateur de la théorie électromagnétique, car ces équations se trouvent pour la première fois établies de façon explicite dans le célèbre

Treatise on Electricity and Magnetism (1873), traduit en France sous le titre de *Traité d'électricité et de magnétisme*. Les conséquences des équations de Maxwell ont été peu à peu dégagées et précisées par une pléiade de physiciens des dernières années du XIX[e] siècle et des premières du XX[e], où se détachent les noms de Lorentz, de Poincaré, de Hertz. La théorie électromagnétique a pris, maintenant qu'on en connaît bien les limites, une perfection définitive, qui s'exprime dans un certain nombre de traités, parmi lesquels on me permettra de citer l'admirable *Electromagnetic Theory* de Julius Adams Stratton (1941), traduit en français sous le titre de *Théorie de l'électromagnétisme*, livre à la fois dur et limpide comme un bloc de cristal.

Sans se dissimuler que la forme mathématique est la seule adéquate, on peut essayer de donner une idée assez précise du contenu physique de ces fameuses équations. Il se compose en fait d'idées très simples, qui résument l'essentiel des expériences de base et en même temps extrapolent leurs résultats avec une audace extrême.

Ces idées, que nous avons déjà pratiquement toutes rencontrées dans les pages précédentes, sont :
— l'indestructibilité de la charge électrique;
— la loi de l'induction;
— la relation (d'Ampère) liant courant et champ magnétique;
— la conservation du flux magnétique.

Mais elles sont exprimées ici d'une façon hautement élaborée et condensée qui leur donne toute leur puissance.

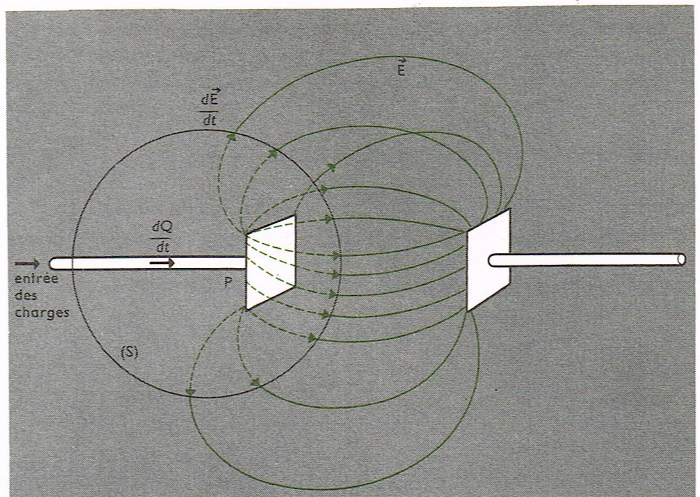
la première loi exprime la conservation de la charge électrique

Ainsi que nous l'avons longuement expliqué, l'invariance de la charge électrique totale est l'un des faits les plus solidement reconnus de la science : une charge positive ne peut disparaître qu'en s'unissant à une charge négative de même grandeur; d'autre part, la charge est la source ou la perte du champ électrique (au sens que nous avons donné à ces termes précédemment). C'est en utilisant les conséquences de ce fait pour approfondir la notion de courant électrique qu'on obtient la première des équations de Maxwell.

Imaginons une région contenant des charges électriques mobiles (ce seront par exemple deux plateaux en regard reliés à une source de tension alternative) et dessinons-y par la pensée une surface fermée fixe S : à chaque instant, le champ électrique a pour sources les charges intérieures à la surface (portées par le plateau P). Cela veut dire que le flux du champ sortant de la surface est proportionnel à la charge intérieure totale.

Supposons, maintenant, cette charge variable : sa variation par seconde (ou, si l'on préfère, sa dérivée par rapport au temps) $\frac{dq}{dt}$ est donc proportionnelle à la variation du flux, ou, si l'on envisage le vecteur « variation du champ » $\frac{d\vec{E}}{dt}$, proportionnelle au flux de ce dernier. Qu'est-ce alors donc que la « variation de charge par seconde » ou le « débit » de charge? Nous l'avons vu : c'est le courant total qui entre dans notre surface. Mais nous avons appris à considérer le courant total comme un « flux de densité de courant ». Si l'on rapproche tout cela, on trouve que le flux de la variation du champ $\frac{dE}{dt}$ est proportionnel au flux de la densité du courant. Cela étant également vrai quelle que soit la surface S (qui n'est là que pour servir en quelque sorte de support au raisonnement), il en résulte une propriété vraie pour les deux vecteurs en chaque point : c'est que leurs divergences sont proportionnelles l'une à l'autre. Ce résultat s'écrit de manière plus complète :

$$\text{div } j = -\epsilon_0 \text{ div } \frac{d\vec{E}}{dt}.$$



Le flux sortant de S vaut : flux $= -\epsilon_0 q$ (le signe — signifie que l'on prend le flux sortant et non le flux entrant). La variation des deux membres avec le temps se traduit par :

$$\frac{d}{dt}(\text{flux de } \vec{E}) = \text{flux}\left(\frac{\delta\vec{E}}{\delta t}\right) = -\epsilon_0 \frac{\delta q}{\delta t} = \epsilon_0 \vec{j}$$

(ϵ_0 flux de densité de courant).

La densité $\vec{j}$ de courant et la dérivée $\frac{dE}{dt}$ du champ ayant même flux (au facteur $-\epsilon_0$ près) à travers toute surface, sont proportionnelles, ainsi que leur divergence. Finalement,

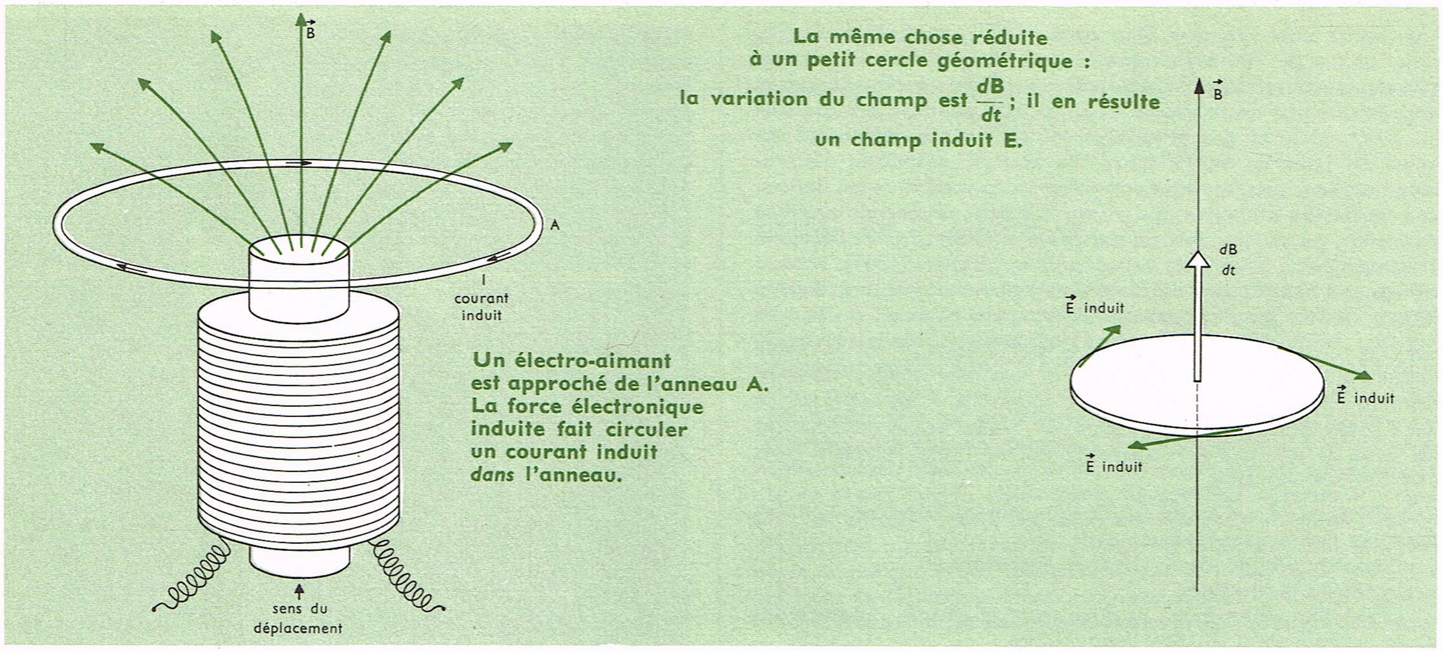
$$\text{div}\left(\vec{j} + \epsilon_0 \frac{\delta\vec{E}}{dt}\right) = 0.$$

C'est donc le vecteur $\vec{j} + \epsilon_0 \frac{\delta\vec{E}}{dt}$ qui est la *densité totale* du courant. C'est un vecteur qui a même structure que le champ magnétique (lignes de force fermées; flux conservatif).

Or, nous avons vu que c'était la valeur nulle de la divergence du courant constant qui exprimait la conservation de l'électricité dans les écoulements permanents. Ce qui précède montre que, pour conserver la même propriété à un courant variable de manière quelconque, il faut ajouter à la densité de courant, au sens ordinaire du terme, le vecteur $\epsilon_0 \frac{dE}{dt}$, qui doit être regardé ainsi comme une véritable densité de courant, bien que ne correspondant à aucun mouvement de charge électrique. On lui conserve, d'après Maxwell, le nom de **courant de déplacement,** bien que ce mot de « déplacement », relatif à des conceptions de Maxwell désormais abandonnées, ait perdu toute signification concrète.

Ce que nous venons d'exposer n'est qu'un calcul et ne nous apporterait que viande creuse s'il ne permettait de formuler nettement et simplement l'hypothèse fondamentale de Maxwell sur l'unicité du courant électrique. En effet, nous admettrons que pour l'observateur extérieur rien d'autre n'entre en ligne de compte que le courant total; ce courant de « déplacement » ne se distingue en rien du courant produit par des charges mobiles, en particulier pour la production des effets magnétiques

A bon droit, on réclamera une preuve expérimentale. Il est bien difficile d'en apporter une directe; un retour à la figure ci-dessus en convaincra aisément : le courant de déplacement est en quelque sorte diffus, tandis que le courant au sens ordinaire est concentré dans le fil d'amenée à la plaque P. Par suite, si nous cherchons l'action produite par une aiguille aimantée,



les participations de l'un et de l'autre courant dans l'effet constaté pourront toujours être mises en discussion. C'est ici qu'il faut savoir se contenter de preuves indirectes, tout aussi convaincantes à la réflexion. **En effet, l'identité des courants dits « de déplacement » et des autres est à la base du mécanisme de la propagation des ondes hertziennes, et la validité de ce modèle est ainsi surabondamment démontrée par ses conséquences.**

la deuxième loi exprime le phénomène de l'induction

En ce qui concerne la loi de l'induction, il faut demander au lecteur un effort particulier pour parcourir presque sans transition le long chemin qui s'étend depuis le rapport bien concret entre l'inducteur et l'induit d'une machine (ou entre les enroulements d'un transformateur) jusqu'à l'abstraction mawxellienne. Essayons-le toutefois. La force électromotrice d'induction prend naissance, avons-nous vu, dans tout circuit à travers lequel le flux magnétique varie. Représentons-nous donc la petite expérience de la figure ci-dessus : un flux magnétique variable est créé dans un anneau de cuivre A par un électro-aimant ; un courant induit prend alors naissance dans l'anneau, ce qui signifie qu'un champ électrique y met en mouvement les électrons. Dépouillons maintenant cette expérience de tout support matériel : l'anneau de cuivre est remplacé par un cercle, simple figure géométrique tracée dans l'espace vide, et ce cercle est si petit que le champ magnétique $\vec{B}$ est constant dans son étendue. Le postulat (ô combien audacieux!) est alors que les choses se passent dans l'expérience idéale essentiellement de la même manière que dans l'expérience réelle. Il n'y a plus de courant puisqu'il n'y a plus rien pour le conduire, mais il reste le champ électrique induit E_i, partout tangent au cercle; quant à la dérivée du flux, nous pouvons la considérer comme le flux de la dérivée du champ magnétique $\frac{d\vec{B}}{dt}$. La question est de savoir comment se traduit dans ce cas idéal la loi de l'induction, appliquée telle qu'on l'a énoncée.

Donnons seulement le résultat sans le démontrer :

Quand le rayon du cercle devient de plus en plus petit, la distribution du champ $\vec{E}$ autour de son centre devient comparable à la distribution des vitesses dans un fluide animé d'un mouvement tourbillonnaire. Si alors on forme la fonction du vecteur $\vec{E}$, qu'on a appelée « rotationnel » ou « tourbillon », la loi de l'induction revient à affirmer l'identité de cette grandeur avec le vecteur $\frac{dB}{dt}$ $\left(\overrightarrow{\text{rot}}\,\vec{E} = \frac{dB}{dt}\right)$.

N'accusons pas le physicien de souffler le chaud et le froid en parlant ici de tourbillon du champ électrique, alors que nous avons insisté sur le fait que le champ électrique statique était non tourbillonnaire; cette propriété n'est vraie, en effet, que pour des champs constants qui ont les charges pour sources, tandis que le champ induit est tourbillonnaire et n'a pas de sources. Dans les cas concrets, le champ peut toujours être décomposé en deux parties : l'une procédant des charges, l'autre des variations du champ magnétique. Mais il est inutile d'effectuer explicitement cette décomposition, car, pour ce qui est de mettre en mouvement les électrons, de produire un courant, il n'y a qu'une seule et unique sorte de champ électrique. On ne demande pas à ses différentes composantes quelle est leur origine, puisque le champ résultant seul compte.

la troisième loi (Ampère) exprime la relation locale entre courant et champ

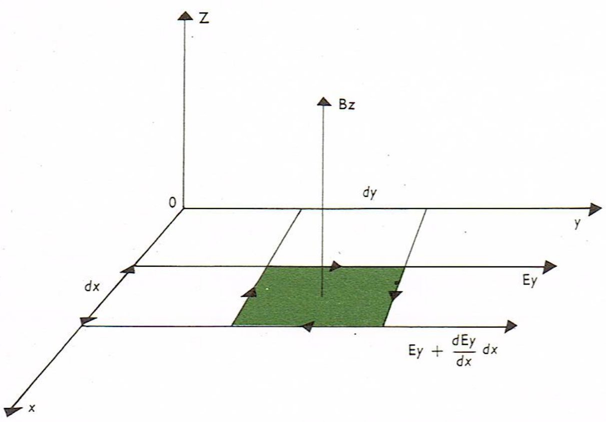
Nous avons déjà énoncé la relation entre le courant et le champ magnétique qu'il produit. Mais le rapprochement avec ce qui précède va nous fournir, sinon une démonstration au sens des mathématiques, du moins une image dans laquelle les deux relations s'éclaireront l'une par l'autre. C'est ici le moment de rapprocher deux schémas déjà étudiés, en cherchant les grandeurs qui s'y correspondent. Le champ $\vec{B}$ (fig. *a*, ci-contre), correspond au champ $\vec{E}$ (fig. *b*). Est-ce maintenant le courant I qui correspond au champ $\frac{dB}{dt}$? Pas tout à fait : c'est le flux du vecteur $\frac{dB}{dt}$ qui intervient pour l'induction, et le courant I est lui-même un flux, celui de la densité de courant. Dès lors, toutes les pièces du puzzle s'assemblent aisément, la relation

champ magnétique ⟶ densité de courant

est la même que la relation

champ électrique ⟶ dérivée du champ magnétique.

L'identité de forme des deux relations n'est pas suffisante cependant pour les expliciter; il faut encore fixer les coefficients de proportionnalité. Rappelons que si le rotationnel



La relation peut se démontrer comme suit, en ne faisant appel qu'à la seule définition élémentaire d'une dérivée partielle.

Au lieu du circuit circulaire, supposons-le rectangulaire, de côtés *dx* et *dy*. La surface vaut *dxdy* et le flux Bz *dxdy*.

Nous devons maintenant évaluer la somme algébrique des produits de champ E par les côtés du rectangle. Les côtés opposés étant parcourus en sens inverse, ces produits, en fait, se retranchent et la différence serait nulle si le champ était constant.

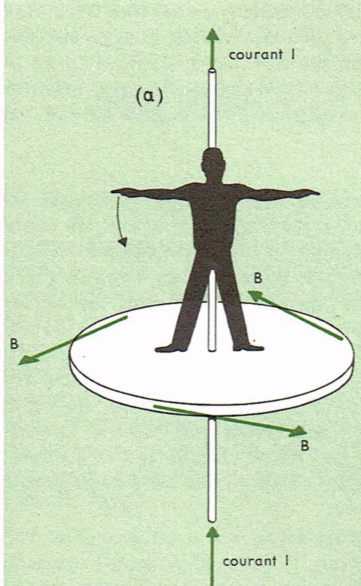
Mais lorsque y croît de dy, Ex croît de $\frac{\delta Ex}{\delta y}$ dy.

Pour le parcours dans le sens indiqué on trouve pour la somme des produits : $\frac{\delta Ex}{\delta y} dy . dx - \frac{\delta Ey}{\delta x} dx . dy$.

Rapprochant ces résultats, on a : $\frac{\delta Ey}{\delta x} - \frac{\delta Ex}{\delta y} = - \frac{\delta By}{\delta t}$.

Comme nous l'avons montré par ailleurs, nous aurons au premier membre la composante suivante Z de rot E. On raisonnerait de même avec les composantes de B suivant *x* ou *y* (petits circuits dans les plans *yoz*, *zox*) ; et par suite, on a :

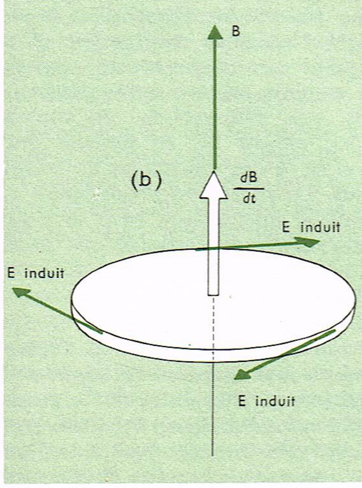
$$\text{rot } E = - \frac{\delta B}{\delta t}.$$



Relation entre un courant rectiligne I et son champ magnétique B

Sur un parcours circulaire quelconque : B = longueur du cercle = μ × I (courant total)

Sur un parcours infiniment petit : tourbillon de B = μ_0 i (densité de courant)



Loi de l'induction appliquée à un circuit infiniment petit :

Sur un parcours infiniment petit : tourbillon de E = $- \frac{dB}{?}$

(le sens — exprime le fait que les champs tournent en sens inverse l'un de l'autre).

de $\vec{E}$ est égal (au signe près) à $\frac{d\vec{B}}{dt}$, celui de $\vec{B}$ est seulement proportionnel à la densité de courant, le rapport de l'un à l'autre étant la « perméabilité » du vide, μ_o.

Ainsi on écrira, dans un symbolisme qui commence à devenir familier,

$$\overrightarrow{\text{rot}}\, \vec{E} = - \frac{d\vec{B}}{dt},$$
$$\overrightarrow{\text{rot}}\, \vec{B} = \mu_o \vec{j}.$$

Remarquons aussi que les sens dans lesquels tournent les champs sont différents d'une figure à l'autre, ce qu'exprime le fait que l'une des équations est écrite avec le signe algébrique —.

Ces expériences mentales, même si le mécanisme en reste déroutant pour le lecteur non spécialisé, comportent de multiples enseignements. **On admirera en passant l'économie de pensée que constitue l'application d'un très petit nombre de procédés mathématiques, toujours les mêmes, à des phénomènes si différents, et la plasticité de ces phénomènes, qui consentent aisément à entrer dans un même cadre.** On sera frappé de la hardiesse et de l'apparente désinvolture de ces extrapolations. Toutes les expériences sur lesquelles se fondent les lois de l'électromagnétisme ont été faites avec des circuits matériels, des bobines, des cadres de galvanomètres, bref, avec des objets à l'échelle humaine. Les équations de Maxwell impliquent tout simplement qu'elles sont applicables à toute échelle, en faisant acte d'absolue confiance dans les procédés de l'analyse mathématique.
De l'échelle des objets, on est passé d'abord à celle de l'infiniment petit; à cet effet, on a commencé par désincarner la loi physique pour lui donner un aspect géométrique, puis on l'a transformée en une relation différentielle établie entre les dérivées des grandeurs prises en un point et à un instant par rapport aux positions et au temps — donc en quelque chose de local, enfermé dans un espace arbitrairement petit, bien plus petit même que l'intérieur d'un atome. La même analyse, qui peut disséquer les grandeurs en leurs dérivées, peut les recombiner en ajoutant de proche en proche un très grand nombre de valeurs — en intégrant des équations sur des étendues arbitrairement grandes. Et nous entendons bien étendre les conséquences de nos équations jusqu'à l'échelle cosmique.

Folie ou tout au moins présomption? Non, méthode rigoureusement justifiée, car elle est moins affirmation qu'interrogation. Les succès éclatants de l'électronique et la maîtrise des ondes



J. C. Maxwell
(1831-1879)



H. A. Lorentz
(1853-1928)

Les équations de Maxwell.

$$\operatorname{dij}\left(\vec{j} + \epsilon_0 \frac{dE}{dt}\right) = 0 \quad \text{conservation du champ électrique}$$

$$\operatorname{rot} E = -\frac{\delta \vec{B}}{\delta t} \quad \text{induction}$$

$$\operatorname{rot} \vec{B} = \mu_0 \vec{j} \quad \text{relation courant-champ}$$

$$\operatorname{div} \vec{B} = \vec{0} \quad \text{conservation du flux magnétique}$$

tout l'électromagnétisme peut se ramener à l'étude de ces formules, très abstraites, mais condensées et générales (ce qui ne veut pas dire que les calculs soient aisés).

Dans le vide la densité de courant se réduira à $\epsilon_0 \frac{\delta E}{\delta t}$ et par suite on a les équations

$$\begin{cases} \operatorname{rot} E = -\dfrac{\delta B}{\delta t} \\ \operatorname{rot} B = -\mu_0 \epsilon_0 \dfrac{\delta E}{\delta t} \end{cases}$$

Si on remplace le symbole rot par les combinaisons de dérivées qu'il représente et que l'on dérive la 1ere équation par rapport au temps, en utilisant la seconde et en se rappelant que div B = O et div E = O (en raison de l'absence de charges) on trouve après quelques calculs simples :

$$\frac{\delta^2 B_x}{\delta x^2} + \frac{\delta^2 B_x}{\delta y^2} + \frac{\delta^2 B_x}{\delta z^2} = \epsilon_0 \mu_0 \frac{\delta^2 B_x}{\delta t^2}$$

et deux équations semblables pour B_y et B_z ce qui s'écrit vectoriellement $\Delta \vec{B} = \epsilon_0 \mu_0 \frac{\delta^2 \vec{B}}{\delta t^2}$, Δ ou « laplacien » désignant la somme des trois dérivées secondes qui se présente ci-dessus.

On trouve les mêmes équations pour le vecteur $\vec{E}$.
C'est cette équation qui *démontre* la propagation sous forme d'onde car elle est identique à celle qui exprime le même phénomène en acoustique et en élasticité, où elle revêt la forme : $\Delta X = \frac{1}{c^2} \frac{\delta^2 X}{\delta t^2}$
où *c* est la *vitesse de propagation* des ondes.

Il résulte de là que la vitesse de propagation des champs électriques et magnétiques vaut :

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}$$



Henri Poincaré
(1854-1912)



Heinrich Hertz
(1857-1894)

électromagnétiques dans la radio et le radar constituent une première et éclatante réponse. L'enrichissement de la pensée scientifique par les théories de la relativité, filles des équations de Maxwell, est une deuxième réponse. Mais, à l'opposé, l'impasse et l'échec radical de cette même théorie électromagnétique devant le comportement des particules élémentaires — et d'abord de son enfant chéri, l'électron — est une troisième réponse, qui s'est révélée encore plus stimulante que les succès ; c'est ce qui apparaîtra en pleine lumière dans le chapitre de la physique corpusculaire, lorsque l'**onde électromagnétique devra non pas céder la place, mais faire une place au photon.**

Revenons ici aux relations locales des champs électrique et magnétique. Dans un espace vide de toute matière et de toute charge, nos équations deviendraient d'une particulière simplicité : le champ électrique, variable, engendre en chaque point un courant auquel s'attache un champ magnétique; les variations de celui-ci engendrent un champ électrique induit, lequel, par le fait même qu'il est variable, constitue un courant. Et la boucle est fermée. Bien loin de conduire au « cercle vicieux » de trop de raisonnements de la vie courante, ce retour au point de départ, exprimé en langage mathématique, donne la propriété fondamentale qu'ont l'un et l'autre champ de s'organiser, lorsque les conditions le permettent, en ondes **se propageant avec une vitesse déterminée.**

vitesse de propagation d'une onde électromagnétique

En attendant de revenir de façon détaillée sur la notion d'onde, indiquons déjà les caractères essentiels de la vitesse de propagation. Elle n'exprime point — comme la vitesse d'un mouvement matériel — le transport d'un objet quelconque d'un point à un autre, mais elle signifie que le phénomène qui se propage (c'est à dessein qu'on emploie un terme très général; cela peut être un mouvement, une déformation, un champ électrique) apparaît d'abord ici, ensuite là, en passant par tous les points intermédiaires, et ce, à des intervalles de temps proportionnels aux éloignements. La vitesse de propagation *c*, commune aux champs électrique et magnétique dans le vide, est reliée simplement à ces grandeurs universelles que nous avons désignées par ε_o et μ_o : $c = \frac{1}{\sqrt{\varepsilon_o \mu_o}}$.

Nous avons remarqué que la valeur numérique de μ_o dépendait des unités choisies; on peut en dire autant d'ε_o. L'idée qui vient alors à l'esprit est qu'il est peu sérieux d'appeler « grandeurs universelles » ces « coefficients » auxquels on peut, par le choix des unités, donner telle valeur numérique que l'on veut. C'est donc maintenant qu'il faut bien prendre garde que

Phot. Bibl. nat., Krenvaldt, X.

tout changement d'unité altérant μ_0 se répercute forcément sur ε_0, et, aussi longtemps que l'on ne change pas les unités de longueur et de temps qui définissent celle de vitesse, le produit $\varepsilon_0 \mu_0$ reste invariable.

En unités M. K. S. A. (en ce qui concerne la vitesse : mètres par seconde), la quantité $\frac{1}{\sqrt{\varepsilon_0 \mu_0}}$ vaut 3.10^8 (ou $2,997\ 9.10^8$ si l'on veut être précis) correspondant à 300 000 kilomètres par seconde en chiffre rond. A ces chiffres, on a reconnu la vitesse de la lumière, et c'est là en effet le résultat essentiel et révolutionnaire de la théorie.
Lorsque Maxwell, vers la fin de son traité, expose pour la première fois la théorie électromagnétique de la lumière, il rapproche les mesures connues à l'époque de la vitesse de la lumière et de l'inverse de $\sqrt{\varepsilon_0 \mu_0}$. Quoique différentes les unes des autres, selon les méthodes employées pour les obtenir, les valeurs trouvées se groupent manifestement. C'est pourquoi il conclut, avec une modestie verbale toute britannique, que sa théorie n'est « certainement pas contredite par la comparaison de résultats tels qu'ils sont ». En effet, elle pouvait passer déjà pour remarquablement confirmée, et le futur devait lui apporter un triomphe éclatant. Par la suite, les mesures ont été multipliées et affinées, et il est frappant de voir les résultats se rapprocher les uns des autres à mesure que progresse la précision. **L'identité des vitesses de propagation est alors, non certes une raison suffisante pour le faire, mais une invitation pressante à identifier lumière et ondes électromagnétiques,** idée dont nous verrons à la fois la fécondité et les limites.

Les équations de Maxwell substituent des actions locales à des forces s'exerçant à distance. Non seulement certaines difficultés de principe sont ainsi levées, mais, encore et surtout, nous pouvons traiter les problèmes dépendant du temps de façon quelconque, sans avoir à considérer d'autre simultanéité que celle d'événements infiniment voisins. Dans l'étude des régimes permanents (où le temps n'intervient pas), des lois de forces s'exerçant à distance peuvent nous satisfaire parce que des questions de simultanéité ou de retard ne se posent pas. A cause de la très grande vitesse des ondes électromagnétiques, on peut encore s'en contenter toutes les fois que les variations sont lentes et les distances petites (à notre échelle). Mais si les champs deviennent très rapidement variables, on ne peut plus se permettre de négliger des retards même très courts : en 10^{-9} seconde, l'onde parcourt 30 cm. L'usage correct des équations de Maxwell (qui n'est pas du tout élémentaire!) est la seule façon satisfaisante d'attaquer ces problèmes.

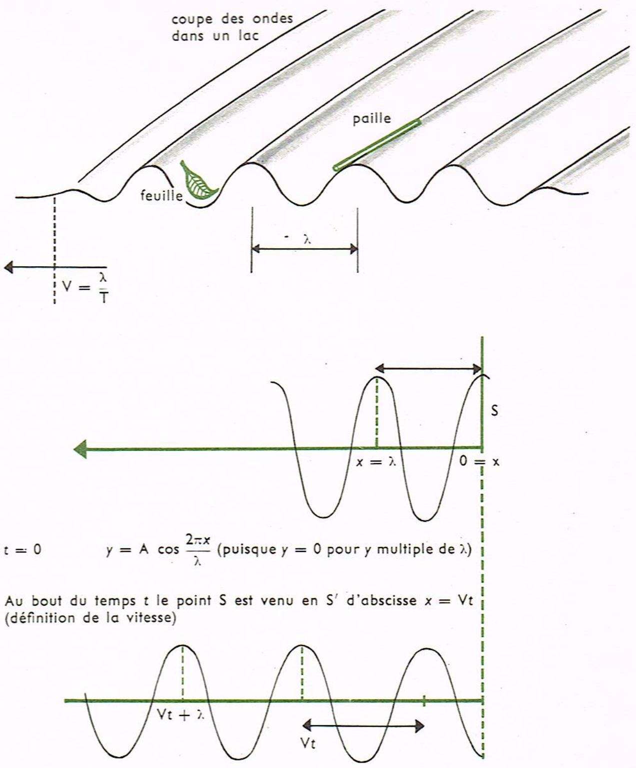
PHÉNOMÈNES ONDULATOIRES ET VÉHICULE ÉLECTROMAGNÉTIQUE

La notion d'onde mécanique est peut-être plus immédiate que celle d'onde électromagnétique. Lorsque notre ancêtre des cavernes s'est approché au bord d'un lac tranquille, il n'a sans doute pas fait que contempler son image, et découvrir l'optique, comme le veut la légende de Narcisse; il a dû aussi faire tomber un caillou, il a dû vouloir toucher du doigt la surface calme de son premier miroir. Alors, il a vu des ronds dans l'eau, des ondes qui s'épanouissaient. Un fétu de paille, une feuille morte se sont alternativement soulevés, puis sont retombés dans chaque creux de l'eau. Mais ils ne se sont pas éloignés de la berge; l'onde ne les a pas emportés. Leurs oscillations ont cessé, et le train d'onde a disparu au loin, en s'amortissant vers le large.

Sur la figure, on a fait une « coupe » pour voir l'onde de profil, comme derrière la vitre d'un gigantesque aquarium. La distance entre deux creux, ou encore entre deux crêtes — plus généralement entre deux points consécutifs à la même hauteur —, c'est la **« longueur d'onde »,** que les physiciens, dans leur amour pour l'alphabet grec, représentent habituellement par la lettre λ (lambda). La distance parcourue pendant une seconde par le front de l'onde, par la perturbation, est la vitesse V de propagation. Le temps mis par le fétu de paille ou la feuille errante pour se retrouver dans deux positions identiques est la période T du mouvement : $\lambda = VT$

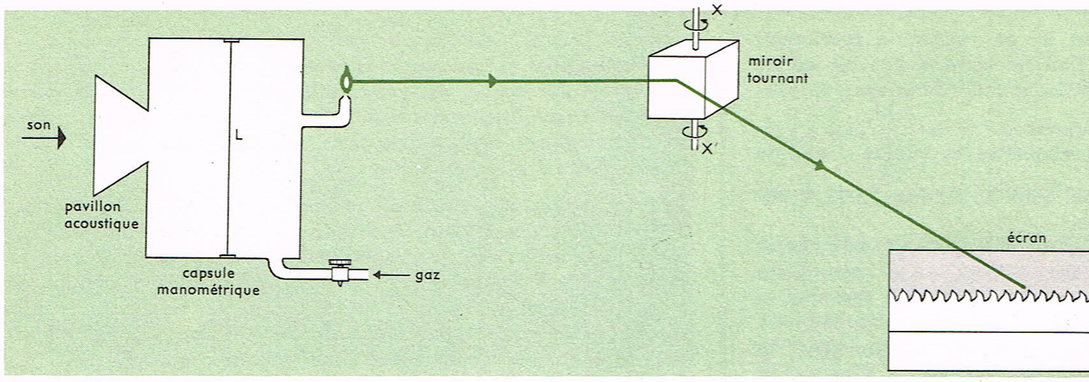
L'amplitude 2A de la vibration est la distance verticale qui sépare les creux des crêtes. Au fur et à mesure que ces ondes s'éloignent de la source, leur rayon augmente continuellement; le front de l'onde s'aplatit, ressemble de plus en plus à une ligne droite (c'est l' « onde plane », que nous allons retrouver).

Dans le mouvement ondulatoire qui vient d'être décrit, l'onde s'appuie sur un milieu matériel, l'eau du lac. Il s'agit d'une vibration mécanique. Ce qui vibre, c'est le milieu de propagation, en l'espèce le liquide. Lors de la propagation du son, le phénomène est tout à fait analogue; l'onde est alors acoustique. L'air vibre, les variations de pression se transmettent jusqu'à notre tympan. Lorsque passe un camion dont le moteur tourne à bas régime, on éprouve une réelle sensation de trépidation correspondant aux basses fréquences acoustiques.



Représentation des ondes.

L'onde a une double périodicité : un déplacement de λ (dans l'espace) ou de T (dans le temps) ne modifie pas l'état de mouvement défini par y. Puisque λ est la période de l'onde dans l'espace et T la période de l'onde dans le temps, à la période T correspond la période dans l'espace VT = λ. On introduit encore la fréquence $\nu = \frac{V}{\lambda}$.



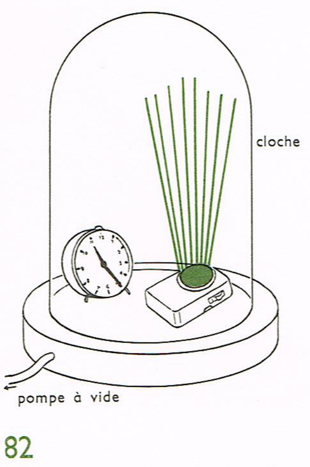
Montage expérimental pour l'enregistrement d'un mouvement périodique, en l'occurrence les sons émis face au pavillon acoustique.

Physiologiquement, c'est encore ce que l'on ressent en écoutant les basses des grandes orgues exécutant une fugue de Bach. La vitre vibre et, quelquefois, se casse lorsqu'un avion a franchi le mur du son au-dessus de la ville. Le stéthoscope permet d'entendre les pulsations, les bruits du cœur.

Une expérience simple et déjà ancienne, qu'on montre toujours dans les lycées, met en évidence la nature ondulatoire du son : un son étant émis devant le dispositif à capsule manométrique, la lame L se met à vibrer, provoquant des changements de pression à l'intérieur de la capsule; le débit de gaz varie avec la pression, et ainsi la longueur de la flamme « suit » la vibration sonore. Un miroir tournant — prisme dont les faces sont des miroirs, et qui est susceptible de tourner autour de l'axe *xx'* — renvoie l'image de cette flamme, l'étale sur un écran disposé à quelque distance et traduit les variations d'amplitude de la flamme sensible, donc du son étudié.

La double périodicité du mouvement vibratoire, dans l'espace et dans le temps, se retrouve dans le phénomène électromagnétique comme dans le phénomène mécanique. Mais comment pousser plus loin l'analogie de ces mouvements vibratoires? Tout d'abord, est-il nécessaire d'imaginer un support élastique pour l'onde électromagnétique? Pour qu'une onde mécanique se propage, il lui faut un « milieu vibratoire. » Une onde sonore est transmise jusqu'à notre oreille, qui la perçoit par l'intermédiaire de toute une « chaîne de transmission » : air ambiant, pavillon de l'oreille, tympan, osselets, etc.; lorsqu'on place une sonnerie à l'intérieur d'une cloche de verre dans laquelle on fait progressivement le vide, le tintement s'évanouit et l'on n'entend finalement plus rien, tandis qu'on voit encore le battant de la sonnette continuer à s'agiter. Une lampe électrique également enfermée dans la cloche continue à briller du même éclat dans le vide le plus poussé. Entre une lointaine étoile et nous la lumière doit bien traverser des régions où il n'y a strictement rien; et pourtant, elle arrive, cette lumière...

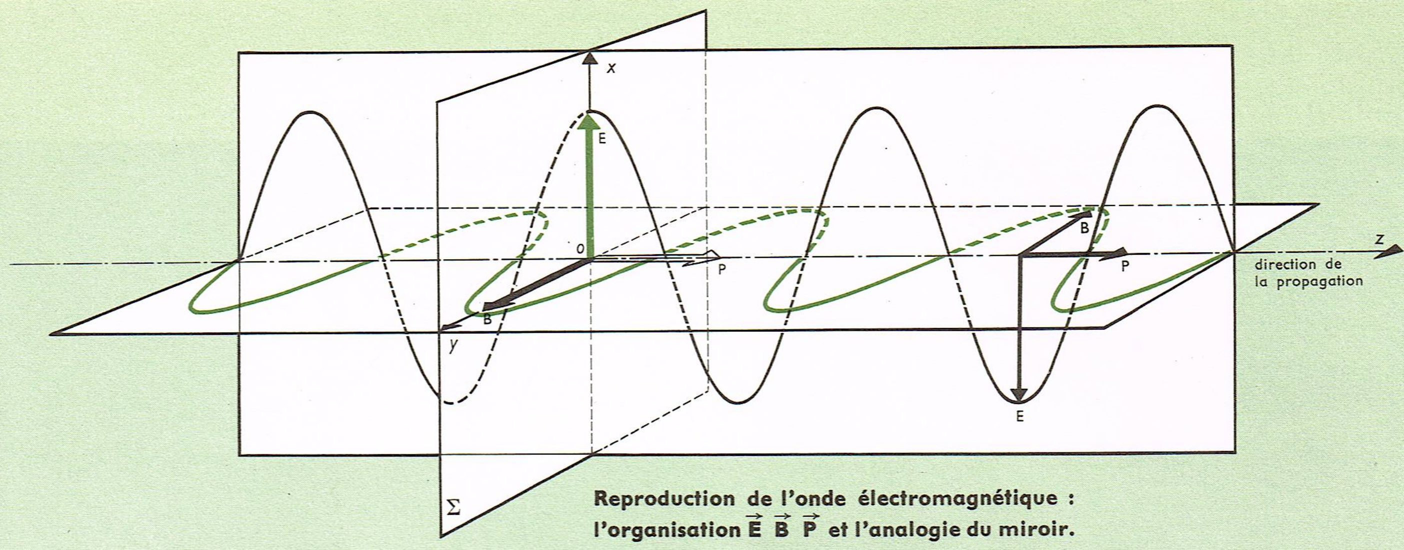
On a tenté, au siècle dernier, pour lui attribuer un support, d'admettre l'existence d'un milieu impondérable, partout répandu, remplissant le vide et imbibant les corps matériels, auquel on avait donné le nom d'**éther** : les phénomènes électromagnétiques ne seraient qu'une manifestation des vibrations périodiques de l'éther. Maxwell, soucieux de donner de ses équations une interprétation mécanique, était favorable à l'existence d'un tel milieu. Pour nous, la propagation des champs électromagnétiques, régie par les équations de Maxwell, de nous oblige pas à introduire une telle notion, source de complications sans fin. Par ailleurs, on n'a jamais réussi à déceler ce milieu. Cependant, il faudra se rappeler que la discussion sur l'existence ou la non-existence de l'éther, polémique passionnée entre mécaniciens et physiciens prolongée sur plusieurs décennies, représente l'un des plus puissants facteurs qui ont conduit Einstein à sa théorie de la relativité.



On continue à voir brûler la lampe, mais on cesse d'entendre sonner le réveil. Le vide transporte les messages lumineux, mais non les messages sonores. Quel est le milieu vibratoire?

Tout au plus, en poursuivant l'analogie de tout à l'heure avec la mécanique et revenant à ce qui se déplaçait périodiquement de bas en haut et de haut en bas (feuille, fétu de paille ou pointe de la flamme), peut-on isoler dans les ondes électromagnétiques une grandeur oscillante, scalaire ou vectorielle, ayant une signification physique déjà connue. En l'espèce, il s'agit incontestablement de **vecteurs,** et l'on se représentera le déplacement périodique de leurs extrémités dans le temps et dans l'espace; ces vecteurs sont précisément les champs électrique $\vec{E}$ et magnétique $\vec{B}$ qui nous sont déjà familiers.

Le type d'onde le plus facile à se représenter est l'**onde plane,** qui a l'intérêt d'offrir une bonne image d'une onde quelconque, pourvu seulement qu'on soit assez loin de l'émetteur et du récepteur : cette onde se propage en ligne droite (suivant ce qu'on appelle le rayon). Et, dans tout plan perpendiculaire à cette direction, les champs qui définissent cette onde sont les mêmes en tout point. On montre que le champ électrique et le champ magnétique sont perpendiculaires l'un à l'autre, et perpendiculaires au rayon; ils sont, comme on dit, en phase, c'est-à-dire passent ensemble par leur valeur maximale, et ensemble changent de sens. On retrouve la longueur d'onde lorsque les distances sont rapportées à l'axe 0z, sur lequel le vecteur $\vec{P}$ caractérise la propagation, en direction et en sens. Cet axe 0z pourrait d'ailleurs aussi être gradué en temps, puisque l'évolution dans l'espace correspond à une évolution dans le temps : à la période d'espace, qui est la longueur d'onde λ, est associée la période de temps T. On reparlera de fréquence, c'est-à-dire du nombre de répétitions, de périodes, par unité de temps. On exprime les fréquences en hertz (Hz), mais il n'est pas rare qu'on le fasse encore en cycles par seconde (c/s). Qui n'a pas entendu annoncer : « Vous êtes à l'écoute de France I, 1 829 mètres, grandes ondes... »? Cela voulait dire que la longueur d'onde, λ, est de 1 829 mètres, ou encore que la fréquence ν est de l'ordre de 164 kilohertz (kHz) : 164 000 fois par seconde environ, les vecteurs $\vec{E}$ et $\vec{B}$ de notre onde (Σ) ont retrouvé la même configuration et la même grandeur. En télévision, la transmission se fait sur des longueurs d'onde voisines de 7 m, c'est-à-dire dans des bandes de fréquences situées entre 40 et 50 millions de cycles par seconde. Les longueurs d'onde de la lumière qui nous vient du Soleil sont inférieures au millième de millimètre, correspondant à des fréquences de l'ordre d'un demi-milliard de millions de cycles par seconde...



Reproduction de l'onde électromagnétique : l'organisation $\vec{E}\ \vec{B}\ \vec{P}$ et l'analogie du miroir.



Les ondes électromagnétiques et notamment les ondes radio se propagent à la vitesse

$$V = c = 300\,000 \text{ km/s} = 3 \times 10^8 \text{ m/s}.$$

ν (fréquence) et λ (longueur d'onde) sont liés par la formule : $\nu\lambda = V = 3\cdot10^8$ m/s

France-Inter émet sur 164 000 hertz :

$$\lambda = 1\,829 \text{ m}, \nu = \frac{3 \times 10^8}{1\,829} = 164\,000$$

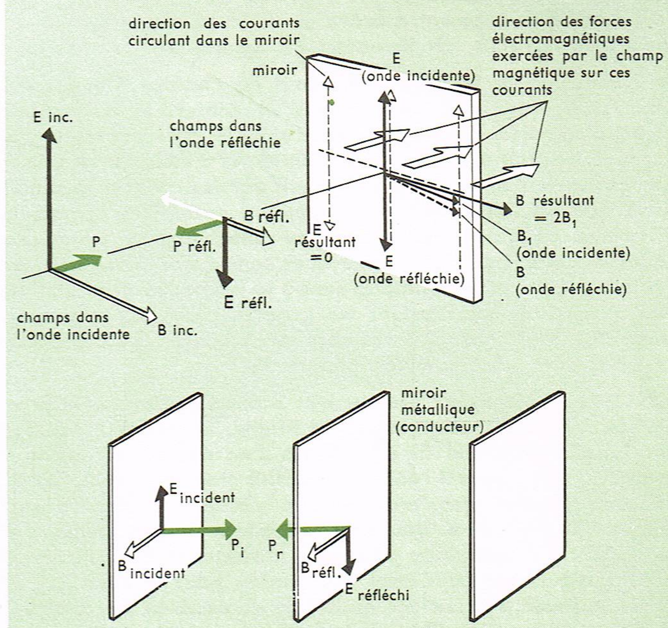
Aux faibles fréquences correspondent les grandes longueurs d'onde.

où se précise l'organisation $(\vec{E}\ \vec{B}\ \vec{P})$ de l'onde électromagnétique

L'organisation $(\vec{E}\ \vec{B}\ \vec{P})$ forme ce qu'on appelle un « trièdre direct », c'est-à-dire qu'un observateur placé à l'intérieur du trièdre et regardant successivement les extrémités des vecteurs $\vec{E}$, $\vec{B}$, $\vec{P}$ devrait tourner la tête dans le sens opposé à celui des aiguilles d'une montre; cette orientation est fixée une fois pour toutes. Supposons qu'une onde plane (Σ incidente) arrive sur un miroir plan de métal poli, par exemple; elle va s'y réfléchir. Le vecteur propagation $\vec{P}$ aura donc, au retour, un sens opposé à celui qu'il avait à l'aller. Quant aux vecteurs $\vec{E}$ et $\vec{B}$, incidents et réfléchis, que vont-ils devenir à la surface du miroir? Le simple bon sens permet de dire que, sur la surface même, ce qu'il y a d'un côté, c'est-à-dire ce qui est réfléchi ajouté à ce qui était incident, est égal à ce qui est transmis. Pour avoir un résultat simple, il faut, comme toujours, idéaliser un peu la situation; nous supposons donc le métal du miroir infiniment conducteur, c'est-à-dire dépourvu de résistance électrique. On voit que, immédiatement en avant du miroir, le champ magnétique est doublé, tandis qu'il est nul de l'autre côté de la face réfléchissante; ce qui a pour conséquence la production d'une nappe de courants superficiels sur le miroir. Le champ magnétique exerce sur ces courants la force de Laplace, d'où une action qui est une pression exercée par l'onde normalement au miroir : la « pression de radiation ».

Phénomène atteignant une très faible intensité, difficile à mettre en évidence expérimentalement, et seulement dans un vide élevé, la pression de radiation n'est pas sans importance, puisqu'elle est, par exemple, responsable de la disposition des queues de comète à l'opposé du Soleil, dont la lumière repousse énergiquement les particules. Phénomène aux conséquences théoriques importantes, en raison du rapport intime entre la pression de radiation et la quantité de mouvement transportée par l'onde électromagnétique, qui elle-même se rattache à l'inertie de l'énergie, comme on le verra dans le chapitre consacré à la relativité.

Phénomène aussi aux conséquences pratiques riches de possibilités. La pression de radiation s'exerce sur les satellites artificiels pendant la partie de leur trajectoire où ils sont éclairés par le Soleil. Les effets, quoique très petits, s'accumulent à chaque révolution des satellites et produisent des déviations sensibles. La même pression servira-t-elle un jour à propulser des vaisseaux cosmiques naviguant au-delà du système solaire?



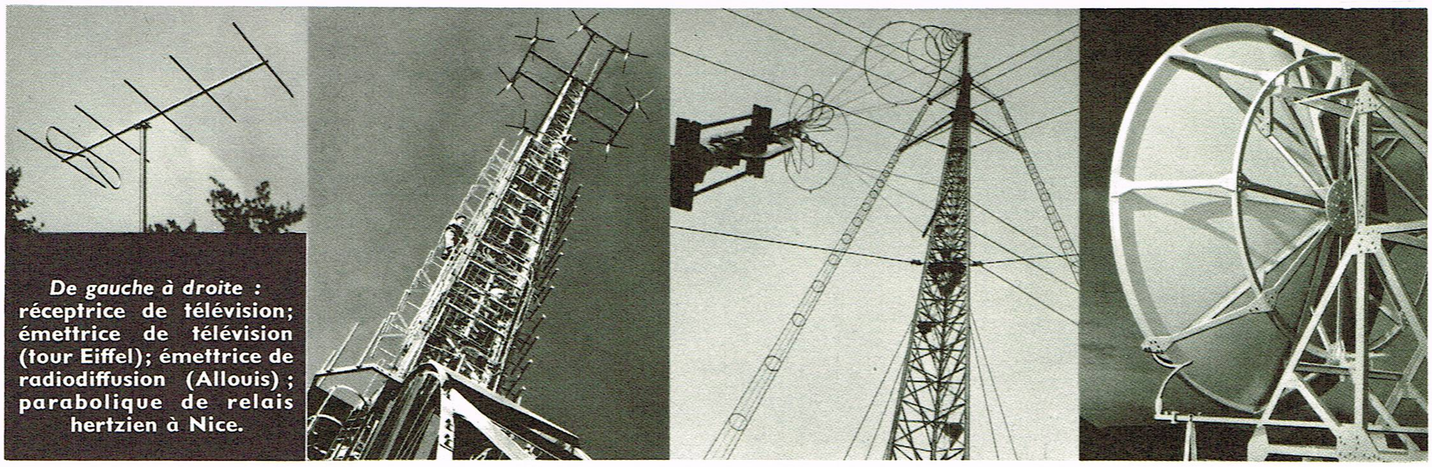
Les trièdres (E_i, B_i, P_i) et (E_r, B_r, P_r) ont le même sens, donc, P_i et P_r étant opposés, on aura dans la disposition indiquée:

$$\vec{E}_t = \vec{E}_i + \vec{E}_r = 0$$

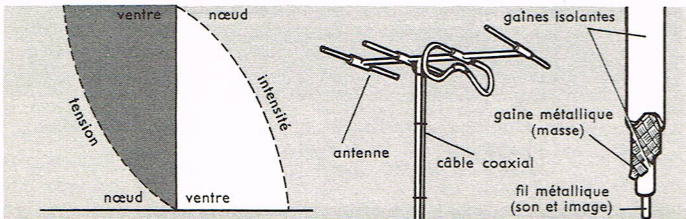
$$\vec{B}_t = \vec{B}_i + \vec{B}_r = 2\,\vec{B}_i$$

Le vecteur annulé est nécessairement $\vec{E}$; le vecteur $\vec{B}$, lui, est doublé.

Réflexion d'une onde plane sur un miroir parfaitement conducteur : mécanisme de la pression de radiation.



De gauche à droite : réceptrice de télévision; émettrice de télévision (tour Eiffel); émettrice de radiodiffusion (Allouis); parabolique de relais hertzien à Nice.



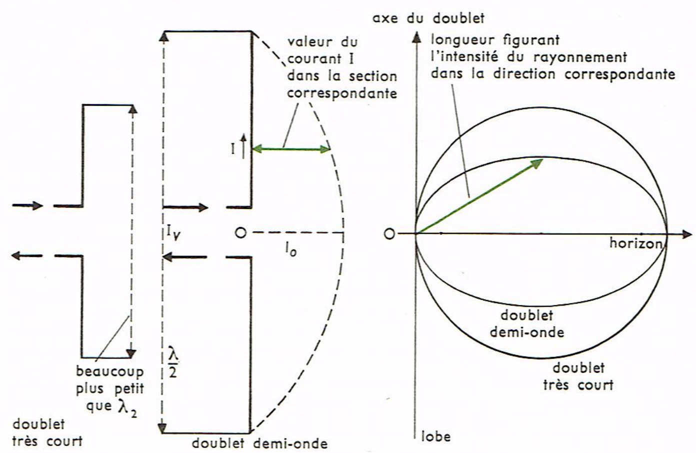
distribution du courant et de la tension dans une antenne — antenne de télévision — descente d'antenne

départ et arrivée des ondes : les antennes

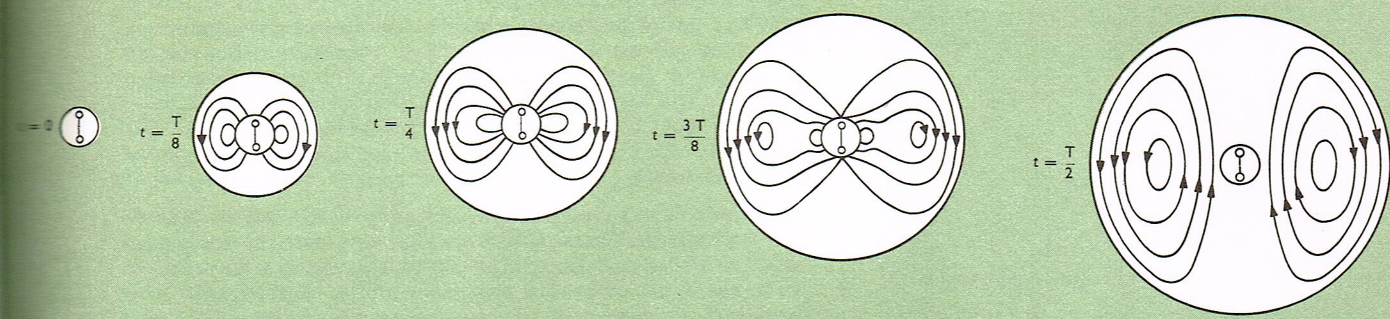
S'il est facile ainsi de voyager par la pensée à cheval sur un rayon lumineux, le départ et l'arrivée de l'onde posent des problèmes d'un tout autre ordre de difficulté. Dans le langage des mathématiques, on demande de déterminer des fonctions de l'espace et du temps représentant les champs $\vec{B}$ et $\vec{E}$ satisfaisant aux équations de Maxwell et conformes aux conditions aux limites, c'est-à-dire, aux valeurs des champs et des courants existant à la surface des antennes d'émission. Celles-ci peuvent revêtir des formes infiniment variées, depuis le simple fil vertical de l'antenne-fouet, jusqu'aux ensembles complexes de cornets, de fentes rayonnantes, de réflecteurs, qui produisent les faisceaux du radar. Il faut avouer sans ambages que les cas concrets même les plus simples sont à la limite des disponibilités de l'analyse, et que, malgré les progrès réalisés, la connaissance des équations différentielles et la construction de leurs solutions sont encore assez imparfaites.

Commençons donc par des cas très simples suffisamment schématisés, c'est-à-dire éloignés de la réalité. Le premier pourrait être l'antenne-doublet (il est difficile d'en donner un exemple concret, tout ce qui est réellement utilisé étant plus compliqué : les plus simples des antennes de réception de télévision comportent déjà deux doublets), alimentée en son milieu par un « câble d'alimentation » (feeder) amenant le courant depuis le générateur. Les conditions aux limites pour le champ, c'est ce qui se passe à la surface de la tige du doublet; par malheur nous n'en savons rien. On ignore donc ces conditions et on considère simplement le courant total déversé par le cable d'alimentation au centre du doublet. Moyennant quoi, on trouve un champ rayonné dont l'amplitude contient des termes en $\frac{1}{r^3}$, en $\frac{1}{r^2}$ et en $\frac{1}{r}$: ces derniers seuls importent à grande distance, et l'on vérifie que l'apparence du champ électrique et du champ magnétique est celle d'une onde plane pourvu qu'on se limite au voisinage du rayon. Il est particulièrement instructif de dessiner la forme des lignes de force du champ électrique en imaginant le doublet émetteur initialement au repos et en faisant des « photographies instantanées » à des instants correspondant à des fractions simples de la période de variation de la charge. Particulièrement digne d'attention est l'image du champ à l'instant où, une demi-période étant écoulée, le doublet est momentanément revenu au repos : alors, il n'y a plus de champ en son voisinage, et celui qui a pris naissance précédemment s'en est complètement détaché. Il vit en quelque sorte pour son compte, et emporte son énergie dans l'espace, détachée de tout support matériel et de toute charge électrique. Le même phénomène se reproduit à la période suivante, et une étude de la figure dispense de plus longs commentaires : la périodicité qui donne naissance à la longueur d'onde s'y voit clairement.

Dans une deuxième approximation, on tient compte de ce fait évident que le courant ne peut pas être le même tout au long de l'antenne. Il est bien obligé, en particulier, de s'annuler à ses extrémités. Avec ces données un peu plus réalistes, on corrige les résultats d'un premier calcul en ce qui concerne l'énergie rayonnée et sa répartition suivant la direction. Si l'antenne est verticale, elle rayonne évidemment de la même manière dans toutes les directions horizontales. Dans un plan vertical, le rayonnement est maximal vers l'horizon, nul au zénith, la décroissance étant lente avec l'élévation; au total, l'antenne est très peu « directive ». A la recherche de nouvelles améliorations, on cessera de considérer l'antenne comme un fil sans dimensions transversales, et on construira une répartition des champs valable à l'intérieur et à l'extérieur de l'antenne et s'étendant hors du conducteur dans tout l'espace en



Antennes doublets et leur rayonnement.



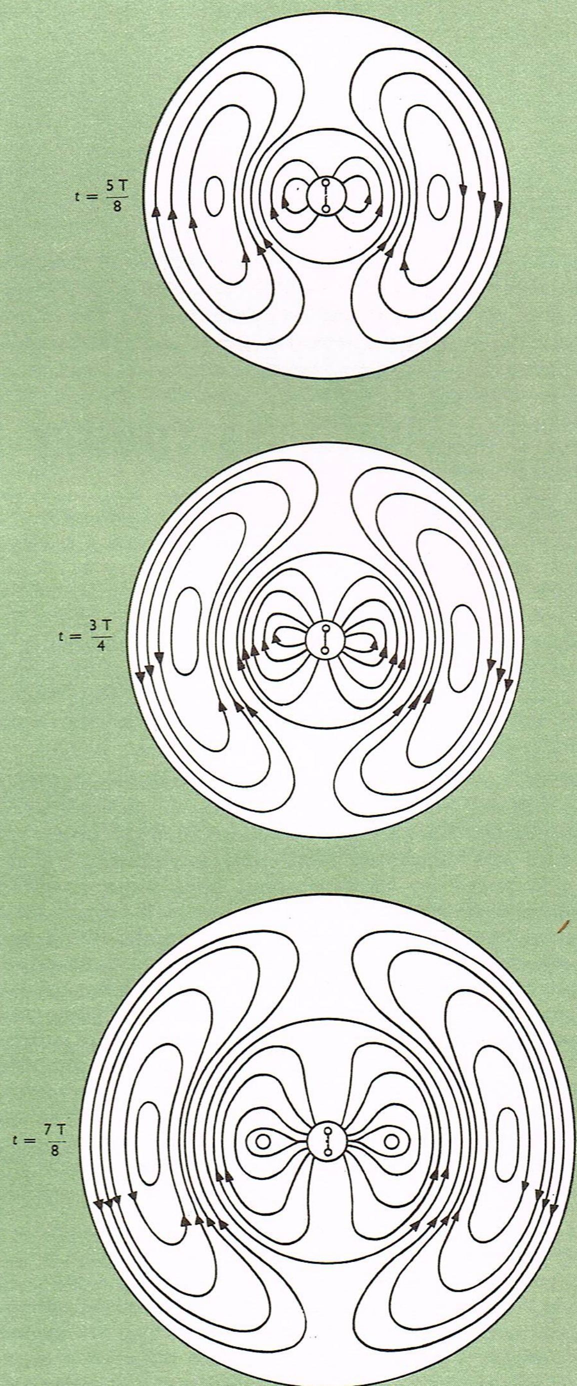
Rayonnement d'un dipôle. Naissance de l'onde électronique.

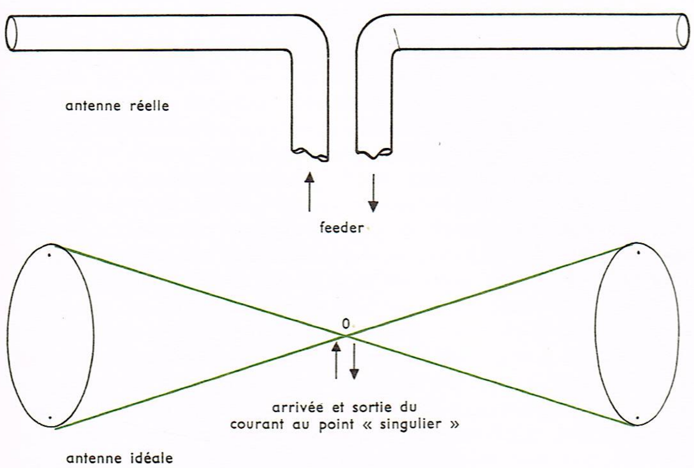
respectant les conditions de passage à la traversée de la surface et celles qu'impose l'alimentation par le feeder. Cela ne sera possible analytiquement qu'en imposant à l'antenne une forme idéalisée, pour le choix de laquelle l'intuition du physicien et la science du mathématicien auront des parts également importantes. Assez curieusement, une des meilleures formes est celle d'un double cône très aigu alimenté par son sommet.

L'emploi de très hautes fréquences correspondant à des ondes très courtes, décimétriques ou centimétriques, a conduit à une manière nouvelle de les transporter depuis les générateurs jusqu'aux antennes. On les fait voyager dans des tubes métalliques de formes diverses dont la paroi intérieure argentée est aussi conductrice que possible; dans ces **guides d'onde,** c'est assez paradoxalement l'espace vide intérieur du tuyau qui « conduit » l'onde, tandis que c'est la paroi conductrice qui l' « isole » de l'extérieur. L'étude des ondes guidées et le guidage lui-même posent un grand nombre de problèmes d'intégration d'équations de Mawxell dont certains atteignent une rare complexité.

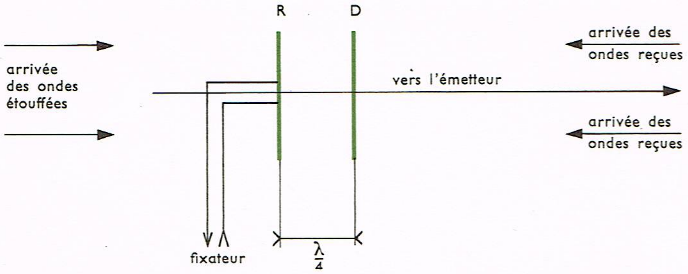
La question de la réception, c'est-à-dire du passage de l'énergie du champ dans une antenne, n'est pas moins intéressante pour l'ingénieur, pour le physicien, et pour chacun de nous dans la mesure où nous sommes auditeurs ou téléspectateurs. Ce qui intéresse l'ingénieur, c'est la recherche de la meilleure utilisation de l'énergie, bien faible à la réception (10^{-5} à 10^{-7} W), mais bien coûteuse à l'émission. Le physicien y trouve des modèles pour les interactions entre lumière et atomes, modèles dont la validité est limitée, c'est entendu, mais dont, jusqu'à nouvel ordre, on ne peut pratiquement pas se passer.

Qu'arrive-t-il donc quand une onde électromagnétique rencontre une antenne? Le champ électrique y fait naître des





Un guide d'onde.



Ce qu'il arrive quand une onde rencontre l'antenne.

D reçoit l'onde 1/4 de période avant le récepteur et réémet dans toutes les directions, notamment dans la direction de propagation : les ondes se renforcent en arrivant sur le récepteur. Au contraire, vis-à-vis des ondes abordant le système dans l'autre sens, l'onde du doublet directeur, qui est maintenant à λ/4 derrière R, prend 1/4 de période de retard en revenant sur le récepteur. Elle se trouve alors 1/2 période en retard sur l'onde incidente, c'est-à-dire en opposition avec elle, avec pour conséquence leur destruction mutuelle. Si le doublet R devient émetteur, l'ensemble émet dans une seule des deux directions : laissons à la sagacité du lecteur le soin de trouver laquelle.

courants qui en reproduisent toutes les variations et constituent le phénomène utile recherché. Mais les choses sont-elles si simples? Dès lors que l'antenne est parcourue par des courants variables, elle se trouve dans les même conditions qu'une antenne émettrice; par suite, elle réémet dans toutes les directions une partie de l'énergie qui l'a frappée, et c'est seulement le surplus qui est disponible pour le récepteur. En poussant les choses à l'extrême, une antenne qui pourrait réémettre tout ce qu'elle reçoit serait complètement inefficace. A l'opposé, l'antenne qui, ayant absorbé, ne réémettrait rien n'est pas concevable, car **la possibilité même d'absorber implique celle d'émettre**.

Il ne faudrait pas croire d'après ce qui précède que la réémission partielle de l'énergie tombant sur une antenne soit seulement un mal nécessaire. De ce phénomène on tire un précieux parti, car il est à la base des propriétés directives dont on a remarqué l'absence quasi totale dans le doublet émetteur décrit précédemment. Il n'y a pas besoin de regarder longuement le paysage des toits de Paris et des grandes villes pour être documenté sur les types variés d'antennes réceptrices de télévision; toutes comportent un doublet relié au récepteur, accompagné d'un nombre variable de doublets isolés qui interviennent uniquement par le rayonnement réémis et renvoyé au doublet actif. Ces ensembles d'antennes sont dirigés vers l'émetteur (si l'on veut capter plusieurs émissions d'origines diverses, il faut autant d'antennes qu'il y a de directions de réception) et ont pour raison d'être de ne recevoir que ce qui vient de lui, en étant dans la mesure du possible sourdes et aveugles aux émissions parasites, et en premier lieu à la réémission des antennes voisines. La théorie des antennes directives est en dehors de l'objet de cet ouvrage, mais des indications en faisant comprendre le principe pourront être données un peu plus loin, lorsque les interférences auront été expliquées.

LA LUMIÈRE ET SES RÉCEPTEURS

L'étude des antennes, représentées schématiquement par le doublet rayonnant, constitue un modèle qui permet d'expliquer l'émission, la transmission et la réception d'ondes électromagnétiques aux fréquences moyennes et relativement élevées. Pourtant, en théorie (dans les techniques de l'hertzien) comme en pratique (pour le physicien étudiant les micro-ondes), ces fréquences sont encore bien basses, comparées aux fréquences optiques. Les longueurs d'onde de la lumière, des couleurs familières qui nous entourent, ne représentent qu'une fraction de millième de millimètre...

Dans cet univers fourmillant d'ondes de l'électromagnétique que nous visiterons bientôt, nous flânerons de manière parfois erratique, nous bornant à poser un regard çà et là, sur quelques régions bien vite délaissées. Mais nous marquerons un arrêt dans le « domaine du visible »; nous nous sentirons retenus par ce véhicule électromagnétique d'un genre à part, l' « onde lumineuse », peut-être parce que toute approche, toute tentative de reconnaissance du monde extérieur, toute description scientifique sembleraient bien timides et incomplètes si elles ne tenaient pas compte des processus physiques et physiologiques par lesquel nous entrons en contact avec ce qui nous entoure, et suivant lesquels nous observons le milieu environnant. L'observation la plus directe est celle qui intéresse, sans intermédiaire, les organes sensoriels. Et chacun sait aussi qu'elle reste toujours associée, explicitement ou non, à ces notions parfois si abstraites que sont le temps et l'espace. Ne paraît-il donc pas naturel de présenter dès maintenant les radiations optiques?

Plusieurs siècles avant notre ère, certainement sans en avoir pris clairement conscience, les Anciens cherchaient déjà à expliquer ces antennes invisibles par lesquelles l'œil pouvait « sentir », « toucher » comme avec la main. Les radiations optiques, agents physiques extérieurs à l'œil, c'est-à-dire les diverses lumières simples, produisent des sensations différentes. A chacune d'elles correspond une couleur différente, bien déterminée par sa longueur d'onde. Pourtant, l'œil le plus subtil parvient quelquefois difficilement à saisir une nuance entre deux teintes très voisines. C'est qu'il n'existe aucune discontinuité entre les radiations du domaine visible, toutes les longueurs d'onde sont susceptibles de se présenter...

Nous venons de parler du domaine visible. C'est par cette gamme restreinte que nous nous introduisons donc dans cet immense ensemble des longueurs d'onde qui constituent le spectre électromagnétique; tout simplement parce qu'elles sont accessibles immédiatement, et de la manière la plus tangible. Enfin parce que le langage le plus simple traduit déjà l'envie de connaître par un « voyons »...

la lumière : une onde électromagnétique

Ne faudrait-il pas, cependant, faire un petit retour en arrière? Nous avons classé la lumière parmi les phénomènes électromagnétiques sans que soient posées quelques questions nécessaires : la nature de la lumière est-elle ondulatoire? s'agit-il d'ondes électromagnétiques? Que pour le physicien la lumière soit avant tout une forme d'énergie transportée d'un endroit à un autre, soit. Mais les questions précédentes ne peuvent être éludées. Ne cherchons pas à imposer comme un acte de foi l'identité du phénomène lumière avec l'ensemble électromagnétique. Il est bien tentant de se dire qu'après tout le pinceau lumineux, le rayon de soleil qui passe à travers le volet ne semblent rien devoir à un phénomène ondulatoire. Et l'on rend compte, en effet, de nombreux résultats en ne faisant intervenir que des rayons, des lignes qui se brisent ou s'incurvent au passage d'un milieu dans un autre. C'est l'optique géométrique. Cela n'est pourtant qu'une première approximation, bien commode certes, mais qui ne permet d'expliquer qu'un nombre restreint de faits, même d'observation courante. Comment, par exemple, la petite lumière anodine qu'on regarde au loin à travers le voile fin des rideaux fait-elle éclater cette

splendide irisation, régulière comme une rosace de cathédrale, dont chacun s'est amusé à suivre les détails, à admirer la finesse des couleurs?

Nous allons voir cependant que le rayonnement lumineux est ondulatoire, plus précisément encore, électromagnétique.

Les équations de Maxwell ont montré que toute variation du champ électrique engendre un champ magnétique variable, et réciproquement. Lors du déplacement de $\vec{E}$ ou de $\vec{B}$, les variations de flux correspondants donnent naissance à ce que nous avons appelé l'« organisation (E,B,P), » c'est-à-dire à l'onde électromagnétique (Σ) se propageant dans la direction P. Lorsqu'un courant de fréquence radio-électrique est envoyé dans une antenne émettrice, tout se passe comme si l'on en secouait les électrons. Une oscillation prend forme. L'onde électromagnétique « se détache » à la vitesse c = 300 000 km/s (c'est aussi la vitesse de la lumière, mesurée directement ou déterminée par différents procédés). L'image qui vient à l'esprit est celle d'une succession de ronds de fumée de forme déterminée qui s'éloignent de l'antenne. Cela est valable au moins pour les longueurs d'onde permettant l'emploi d'une antenne dont l'allure est familière et, sans revenir sur ce qui a été exposé dans les pages précédentes, on conçoit aisément qu'un phénomène identique se produise sur l'antenne de réception.

Les électrons secoués par l'onde incidente traduisent leur agitation ordonnée par un signal sur le récepteur, qui a été prévu pour détecter et amplifier les signaux reçus et les réarranger de façon intelligible. L'onde étant caractérisée par sa longueur d'onde λ (ou par sa fréquence ν), on sait très bien que la réception n'est possible que si l'on a accordé son récepteur sur l'émetteur choisi, c'est-à-dire si on l'a placé dans les conditions de travail correspondant à celles de l'émetteur (ν à recevoir sensiblement égale à ν émise).

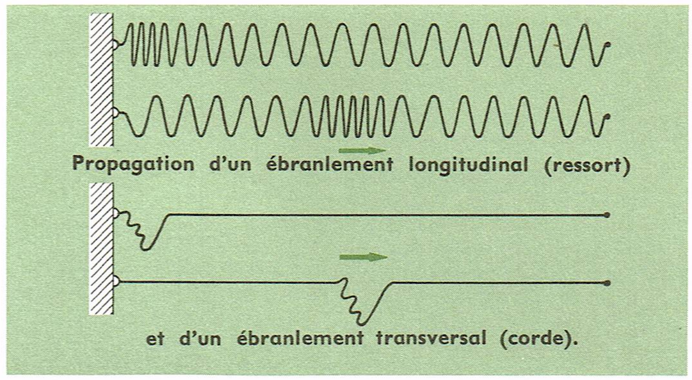
A peine différemment peut-on se représenter l'émission de rayons X : dans un tube approprié, le bombardement de l'anticathode par des électrons arrivant à vitesse convenable fait jaillir un faisceau de rayons X dans la région de rapide décélération qui est celle de la cible. La vitesse de propagation du rayonnement X est mesurable : elle est de $c = 3.10^8$ m/s comme celle de la lumière.

Autre fait expérimental : l'accélération de charges élémentaires dans l'un des grands appareils décrits au début de cet ouvrage — un synchrotron par exemple, dans lequel on fait tourner des électrons de plus en plus vite — met en jeu une énergie électromagnétique importante, limitée par le seul fait que, lorsque la vitesse des électrons se rapproche de c, l'énergie électromagnétique rayonnée par les électrons compense tout apport nouveau. Or, il y a émission de lumière, observable au voisinage de la trajectoire. Cette lumière est bien d'origine électromagnétique.

Il existe d'autres preuves, plus anciennes et non moins convaincantes de la nature électromagnétique de la lumière, de cette unité du rayonnement électromagnétique où l'optique apparaît comme un maillon entre les grandeurs à l'échelle humaine et l'infiniment petit. Nous les retrouverons plus tard, en côtoyant plus intimement le vecteur lumière (vecteur de Fresnel) dans l'explication de phénomènes particuliers : la polarisation de la lumière, les interférences, la diffraction.

qu'est-ce que le vecteur lumière?

Si la nature de la lumière ne diffère pas de celle des ondes hertziennes ou du rayonnement X, si, d'un bout à l'autre du spectre, les différences essentielles se situent au niveau des émetteurs, des détecteurs, de leur technologie, faisons donc ici plus ample connaissance avec ce représentant des ondes électromagnétiques : le vecteur lumière.



Régie par les équations de Maxwell, la grandeur lumineuse est vectorielle. Il nous faut la situer de manière plus précise. Est-elle de nature magnétique? électrique? Ce quelque chose qui vibre reste-t-il attaché à la direction de propagation? S'agite-t-il en tous sens dans le plan de l'onde?

Revenons donc un instant à la figure représentant les ronds qui se dilatent sur l'eau du lac tranquille, et regardons d'un peu plus près les rides de la surface un moment troublée par la chute d'une pierre. La feuille qui se balance — sans s'éloigner du bord — se déplace pourtant quelque peu d'avant en arrière, dans un mouvement de va-et-vient qui s'amortit, lui aussi, quand l'amplitude verticale décroît.

Ce mouvement bien particulier de la feuille, c'est celui même d'un ressort à boudin dont on aurait pincé brusquement quelques spires, aussitôt relâchées. On verrait se déplacer, le long du ressort, un ébranlement d'un type nouveau : les spires sembleraient se resserrer dans certaines régions — zones de contraction ou de compression —, tandis qu'elles se dilateraient dans d'autres — zones d'expansion. Ainsi, une spire particulière effectuerait de petits allers et retours périodiques : on aurait affaire à une **onde longitudinale.**

Prenons maintenant l'extrémité libre d'une corde attachée à l'autre bout à un point fixe et frappons un coup sec sur cette corde, au moyen d'un bâton par exemple. L'ébranlement se propage vers le mur, chaque point de la corde passant alternativement par une position haute et basse. L'onde (bien vite amortie dans cette expérience) est **transversale.**

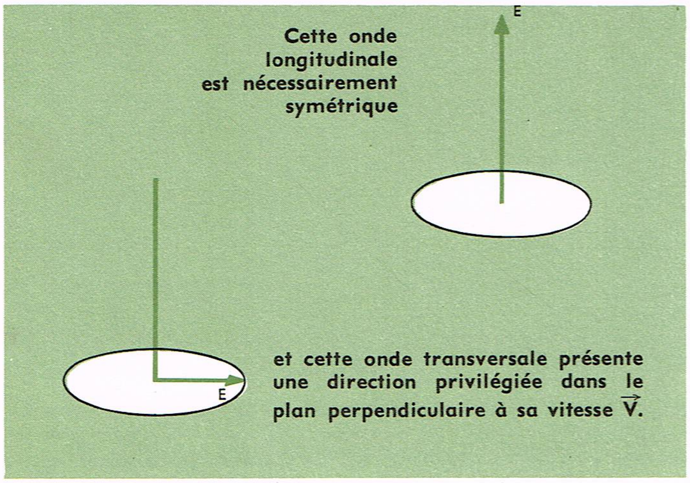
Sur l'eau, le mouvement des vagues devient complexe, à la fois transversal et longitudinal. Le baigneur se sent monter et descendre, mais il est aussi emporté au large puis rejeté vers la rive. L'impression est particulièrement nette lorsqu'on a tout juste pied au bord d'une plage quand la mer commence à s'agiter.

Les vibrations dans les solides mettent en jeu des ondes longitudinales du type de celles du ressort à boudin, et des ondes transversales, qui agitent les molécules ou les atomes comme la feuille du lac.

longitudinal ou transversal?

De quel type est la vibration lumineuse? Longitudinal ou transversal?

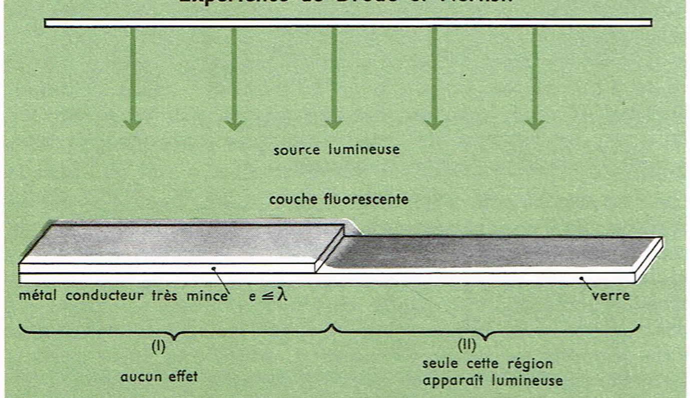
Il est facile de répondre à une telle question : une vibration longitudinale présente nécessairement une symétrie de révolution autour de son axe de propagation, et toutes les directions du plan perpendiculaire jouent le même rôle, par raison de symétrie. Au contraire, le plan perpendiculaire à l'axe de propagation d'une vibration transversale possède, non moins nécessairement, une direction privilégiée unique. Et des phénomènes doivent en témoigner; la polarisation (que nous verrons plus loin) en est un bel exemple : **nous pouvons donc affirmer le caractère transversal de l'onde lumineuse.**



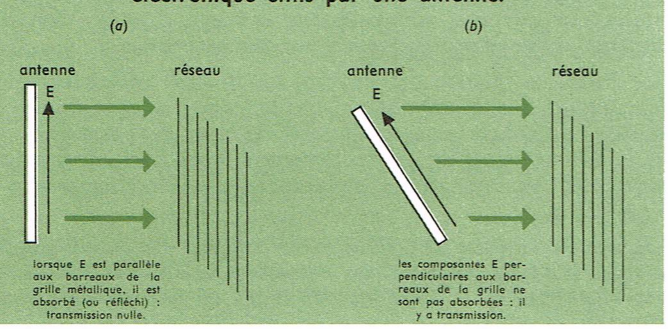
Cette onde longitudinale est nécessairement symétrique

et cette onde transversale présente une direction privilégiée dans le plan perpendiculaire à sa vitesse $\vec{V}$.

Expérience de Drude et Nernst.

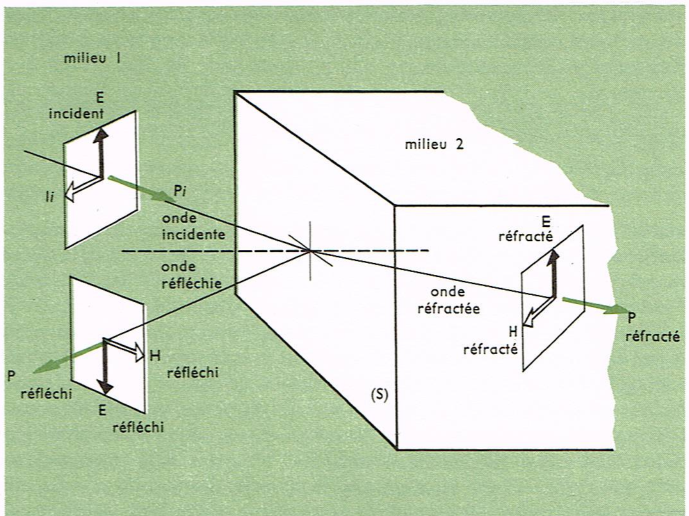


Effet d'une « grille » (réseau) sur le rayonnement électronique émis par une antenne.



Nous savons de plus qu'elle est de nature électromagnétique; le vecteur (baptisé « vecteur de Fresnel ») qui la représente, quel est-il? est-ce le vecteur électrique $\vec{E}$ ou le vecteur magnétique $\vec{B}$?

Le raisonnement suivi pour la réflexion d'une onde plane arrivant normalement sur un miroir infiniment conducteur peut être exactement calqué dans le cas d'une incidence quelconque et pour une onde lumineuse. Par ailleurs, différentes expériences (en ondes stationnaires par exemple) montrent que si le miroir est métallique, l'énergie lumineuse détectable dans le plan du miroir est minimale (expérience de Drude et Nernst). Ainsi, le vecteur de Fresnel est donc identifiable au vecteur champ électrique $\vec{E}$. En effet, si la résultante des champs $\vec{E}$ (incident et réfléchi) est nulle sur le miroir, l'amplitude et l'énergie correspondante seront nulles.



Les conditions aux limites sont celles qui sont définies sur la surface de séparation (S) des deux milieux, surface sur laquelle se produit une réflexion, une réfraction — ou les deux; elles s'expriment par le fait que les résultats des composantes tangentielles de E et H doivent conserver la même valeur en des points voisins situés de part et d'autre de la surface (S).

le rayonnement d'une antenne et la lumière : une seule explication

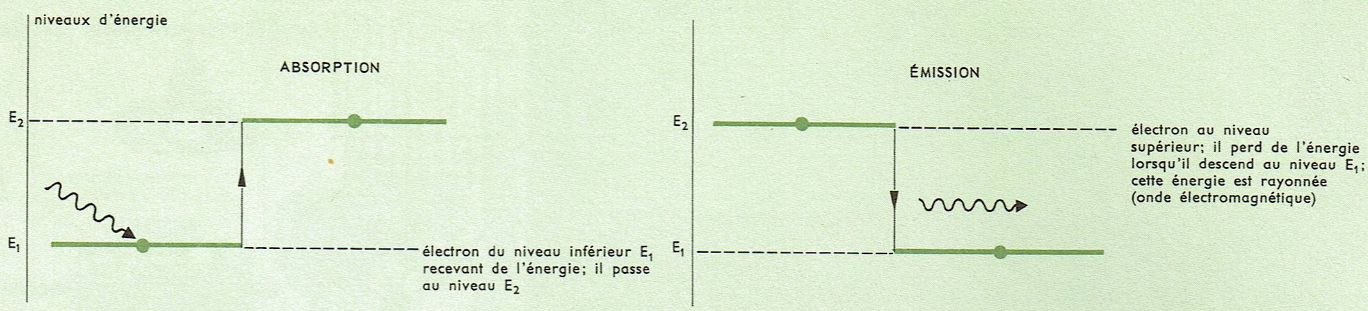
Un phénomène analogue se produit dans le domaine des très petites ondes. Les éléments des miroirs paraboliques métalliques qui entourent l'antenne d'émission, ou, réciproquement, qui jouent le rôle de collecteur dans un radiotélescope et concentrent le rayonnement sur l'antenne de détection placée au foyer, doivent être convenablement orientés par rapport à l'antenne.

La définition du plan d'onde (Σ) n'implique pas une orientation particulière de $\vec{E}$ et $\vec{B}$, pourvu que ces vecteurs restent perpendiculaires et que leur plan soit perpendiculaire au vecteur de propagation $\vec{P}$. Donc $\vec{E}$, c'est-à-dire le vecteur de Fresnel, peut prendre toutes les orientations possibles dans ce plan; $\vec{B}$ s'en déduira.

Dans le domaine des ondes hertziennes, le vecteur $\vec{E}$ possède souvent une orientation privilégiée, mise en évidence dans l'expérience schématisée ci-contre : une série de fils conducteurs parallèles (réseau) étant disposés devant l'antenne rayonnante, il y a réflexion de l'onde électromagnétique, avec un minimum de transmission lorsque celle-ci est elle-même parallèle au réseau. La situation s'inverse lorsque l'antenne est perpendiculaire au réseau : réflexion minimale et transmission maximale. L'antenne effectue donc un choix parmi les orientations possibles de $\vec{E}$; elle rayonne une onde dont le vecteur champ électrique est parallèle au doublet qu'elle représente. Ce triage rend compte d'un phénomène nouveau : la polarisation du vecteur $\vec{E}$. Nous verrons quel parti savent en tirer les opticiens.

quelle est la source du rayonnement ?

Sans doute est-il apparu, au long de ce chapitre, une ombre de mystère planant sur la nature intime des choses. On a bien admis les ondes, tels des ronds de fumée s'échappant de l'antenne. On a bien voulu imaginer des électrons secoués par les



Mécanisme de l'absorption et de l'émission de rayonnements électromagnétiques.

impulsions fournies par l'émetteur. On a accepté cette remarque, suggérée par le mouvement des charges donnant naissance à l'onde électromagnétique, qu'à un certain retard près, dû au fait que la propagation n'est pas instantanée, l'onde détectée traduisait un état vibratoire de la source (en optique comme en hertzien). Mais là, justement, est le fond du problème. Comment l'onde lumineuse est-elle produite? D'ou vient la lumière, perçue par l'œil ou noircissant le film photographique? Quelle est son origine physique précise?

Voici venir le moment où l'électromagnétisme ne semble plus tout expliquer de manière entièrement satisfaisante...

Sans anticiper sur la suite, on peut fort bien encore se représenter le grain presque ultime de matière, l'atome, comme un petit système solaire, pourvu de ses satellites, les électrons, qui gravitent autour de son centre, le noyau. Lorsque, pour une raison ou pour une autre, un électron est bousculé hors de son orbite et précipité sur une autre orbite, il va se produire un échange d'énergie. **Cette énergie est électromagnétique.** Un ensemble (E, B, P,) se constitue : il y a rayonnement électromagnétique. Au saut d'un électron correspond une onde de fréquence bien déterminée, fonction de l'énergie échangée. Cette onde peut d'ailleurs aussi bien être absorbée qu'émise. Parler d'absorption revient à dire qu'une certaine quantité d'énergie a été fournie à l'électron, qui l'a utilisée pour changer d'orbite. Réciproquement, un ensemble atomique peut fournir, spontanément ou non, un paquet d'ondes électromagnétiques dont le tableau des fréquences permet de le décrire, de le caractériser. Mais limitons-nous à cette remarque, pour prometteuses qu'apparaissent ces considérations. Elles nous conduiraient à envisager d'autres concepts, en particulier celui de particules de rayonnement. Et ce serait largement déborder le cadre de ce chapitre.

UN SEUL SPECTRE, MAIS PLUSIEURS RÉCEPTEURS

Le physicien étudie les tempêtes magnétiques, il mesure l'intensité des champs au voisinage des taches solaires. Sur son traîneau lourdement chargé d'appareils, le moderne explorateur des glaces polaires observe et mesure, photographie les aurores boréales. L'astrophysicien scrute des mondes si éloignés de nous que notre Terre n'existait pas encore lorsqu'ils nous envoyèrent ces signaux aujourd'hui reçus, enregistrés dans la grande station permanente d'un observatoire. L'œil rivé au bout de son télescope, l'astronome guette patiemment, nuit après nuit, le passage d'une étoile. Le radio-astronome « écoute » les bruits venant d'une région obscure du ciel, scrute l'atmosphère des planètes, détermine l'époque d'un ancien cataclysme cosmique. Lorsque « Lunik » est passé derrière la Lune, ce petit satellite a pris des clichés; il les a même retransmis vers la Terre, tout comme les postes de radio acheminent et regroupent informations et messages.

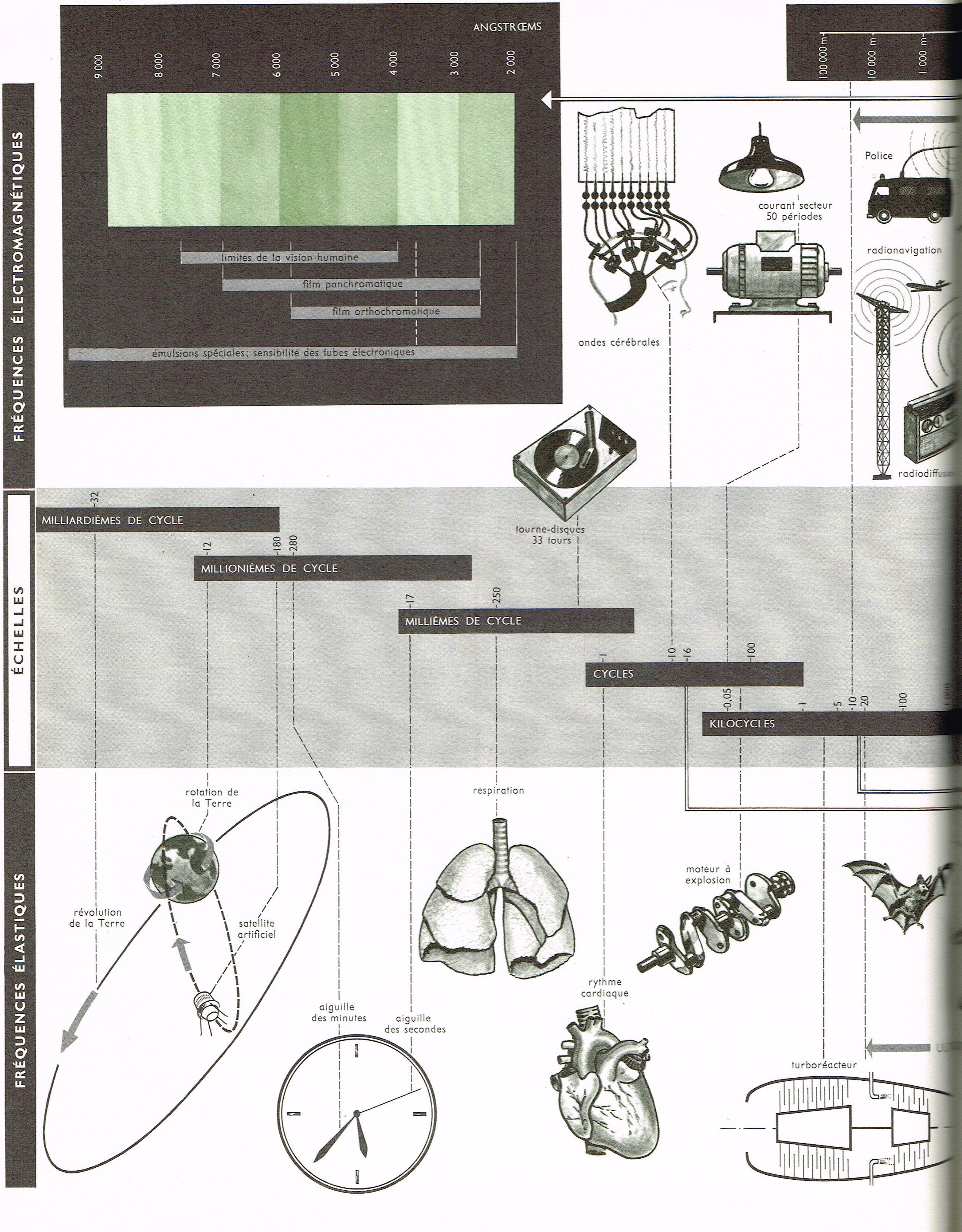
Dans tous ces cas, d'apparence si divers, un même transmetteur de renseignements, un même messager a été utilisé. Ces données si diverses du monde extérieur ont été transportées au moyen d'une même catégorie de véhicules. **Ces véhicules sont les ondes électromagnétiques.**

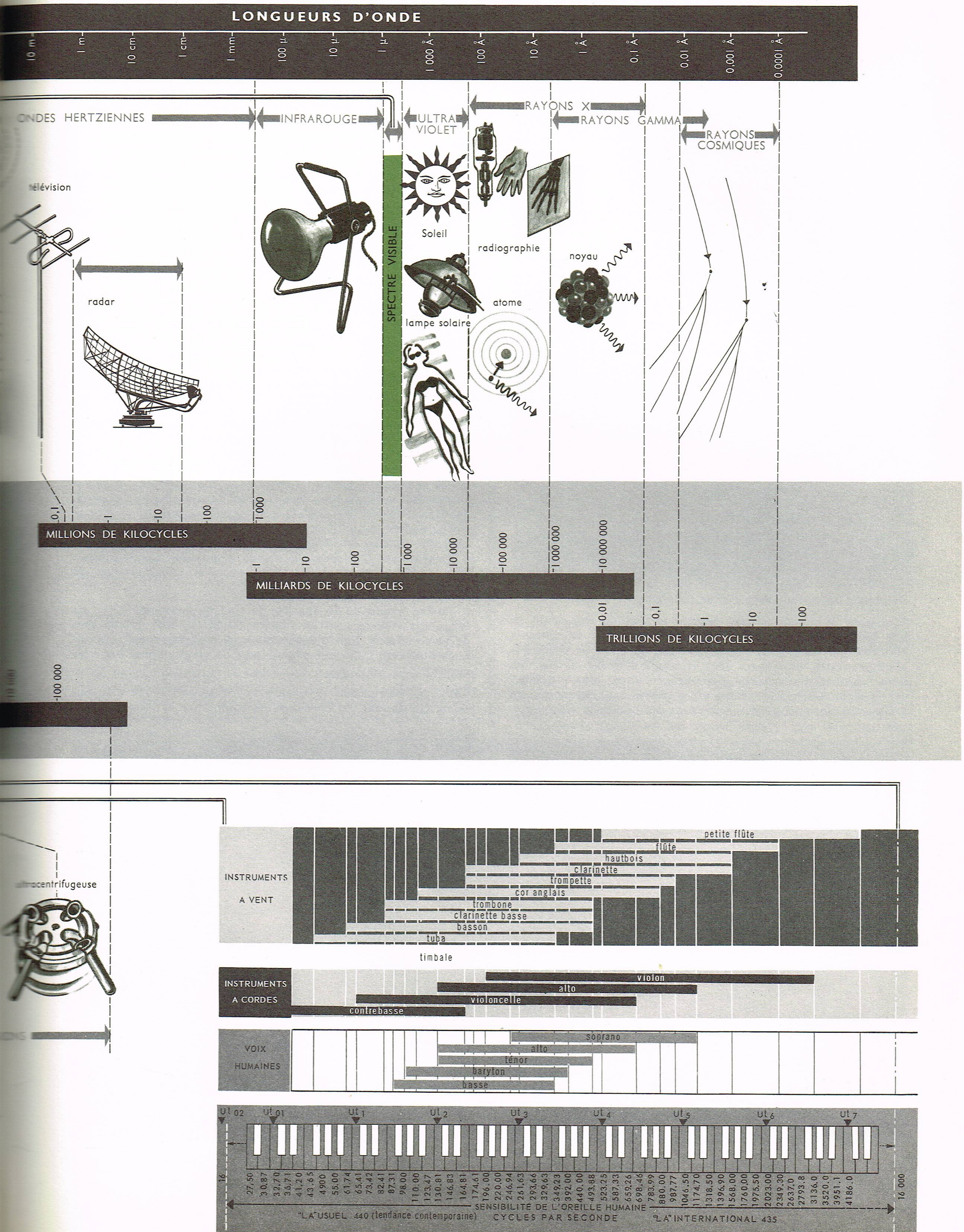
Cependant, les outils des savants, des techniciens, de « Lunik », apparaissent bien différents les uns des autres. Y a-t-il une commune mesure entre le télescope géant et le minuscule émetteur-récepteur installé sur le satellite? En quoi diffèrent essentiellement les transmetteurs, les véhicules? Si sa nature est la même lorsqu'elle apporte à l'astronome le message de l'étoile ou lorsqu'elle renvoie sur Terre, par le truchement de signaux radio-électriques codés, les fragments d'images glanés derrière la Lune, l'onde électromagnétique présente au moins un caractère variable dans l'un et l'autre cas : sa longueur d'onde (1).

diversité des longueurs d'onde

Compte tenu des idées et des résultats de l'électromagnétisme, nous allons essayer de cheminer avec ces véhicules électromagnétiques et de les interpréter; et ce depuis les grandes longueurs d'onde jusqu'aux plus petites (c'est-à-dire jusqu'aux fréquences très élevées), réalisant ainsi l'unité du spectre électromagnétique. Nous en profiterons pour jeter un coup d'œil à travers les fenêtres d'exploration du monde que nous ouvrent les gammes de longueurs d'onde utilisables par nos appareils. Elles sont étroites, réduites par des phénomènes dont nous ne nous sommes pas encore rendus maîtres. Et pourtant, l'instrument de la vision, le sens majeur, l'œil, ne représente plus qu'un récepteur très limité devant la diversité des radiations explorables. Chaque domaine de longueurs d'onde met en œuvre des techniques qui lui sont propres, nécessite des instrumentations particulières dont les possibilités s'étendent tout au plus sur quelques

(1) Pour les petites longueurs d'onde, on utilise le plus souvent les unités suivantes : le micron (μ), qui vaut un millième de millimètre; l'angström (Å), un dix-millième de micron; l'unité X (X), un millième d'angström; le fermi (F), un centième d'unité X ($1\ \mu = 10^{-4}$ cm; 1 Å $= 10^{-4}\ \mu = 10^{-8}$ cm; 1 X $= 10^{-3}$ Å $= 10^{-11}$ cm; 1 F $= 10^{-13}$ cm). Les fréquences sont données en hertz (Hz), kilohertz (kHz), mégahertz (MHz) (1 kHz $= 10^3$ Hz; 1 MHz $= 10^6$ Hz). Rappelons la formule $\lambda\nu = c$ (où $c = 3.10^8\ m = 3.10^{10}$ cm $= 3.10^{14}\ \mu = 3.10^{18}$ Å $= 3.10^{23}$ F, et où ν est exprimé en hertz).


FRÉQUENCES ÉLECTROMAGNÉTIQUES
ANGSTRŒMS
9 000
8 000
7 000
6 000
5 000
4 000
3 000
2 000
limites de la vision humaine
film panchromatique
film orthochromatique
émulsions spéciales; sensibilité des tubes électroniques
100 000 m
10 000 m
1 000 m
Police
courant secteur 50 périodes
radionavigation
ondes cérébrales
tourne-disques 33 tours
ÉCHELLES
MILLIARDIÈMES DE CYCLE
32
MILLIONIÈMES DE CYCLE
12
180
280
MILLIÈMES DE CYCLE
17
250
CYCLES
1
10
16
100
KILOCYCLES
0,05
1
5
10
20
100
FRÉQUENCES ÉLASTIQUES
rotation de la Terre
révolution de la Terre
satellite artificiel
respiration
moteur à explosion
rythme cardiaque
aiguille des minutes
aiguille des secondes
turboréacteur



LONGUEURS D'ONDE
10 m
1 m
10 cm
1 cm
1 mm
100 µ
10 µ
1 µ
1 000 Å
100 Å
10 Å
1 Å
0,1 Å
0,01 Å
0,001 Å
0,0001 Å
ONDES HERTZIENNES
INFRAROUGE
ULTRA VIOLET
RAYONS X
RAYONS GAMMA
RAYONS COSMIQUES
télévision
radar
SPECTRE VISIBLE
Soleil
lampe solaire
radiographie
atome
noyau
0,1
1
10
100
1 000
MILLIONS DE KILOCYCLES
1
10
100
1 000
10 000
100 000
1 000 000
10 000 000
MILLIARDS DE KILOCYCLES
0,01
0,1
1
10
100
TRILLIONS DE KILOCYCLES
100 000
ultracentrifugeuse
INSTRUMENTS A VENT
petite flûte
flûte
hautbois
clarinette
trompette
cor anglais
trombone
clarinette basse
basson
tuba
timbale
INSTRUMENTS A CORDES
violon
alto
violoncelle
contrebasse
VOIX HUMAINES
soprano
alto
ténor
baryton
basse
Ut 02
Ut 01
Ut 1
Ut 2
Ut 3
Ut 4
Ut 5
Ut 6
Ut 7
16
27,50
30,87
32,70
36,71
41,20
43,65
4900
55,00
61,74
65,41
73,42
82,41
87,31
98,00
110,00
123,47
130,81
146,83
164,81
174,61
196,00
220,00
246,94
261,63
293,66
329,63
349,23
392,00
440,00
493,88
523,25
587,33
659,26
698,46
783,99
880,00
987,77
1046,50
1174,70
1318,50
1396,90
1568,00
1760,00
1975,50
2023,00
2349,30
2637,0
2793,8
3136,0
3520,0
3951,1
4186,0
16 000
SENSIBILITÉ DE L'OREILLE HUMAINE
"LA" USUEL 440 (tendance contemporaine)
CYCLES PAR SECONDE
"LA" INTERNATIONAL 435

octaves (lorsque la fréquence d'un son double, on dit qu'on monte d'une octave, en empruntant à la musique un terme consacré). Ce n'est que très récemment, depuis seulement quelques années, que l'on a réussi à faire se recouvrir, sur de minces bandes, les plages d'utilisation de deux catégories d'appareillages spécifiques de domaines juxtaposés (hertzien et optique, par exemple). Nous entrerons ensuite un peu plus avant dans l'optique, par le biais de son aspect ondulatoire, celui pour lequel l'outil privilégié de raisonnement est le modèle électromagnétique; quelques éléments de la polarisation en optique se présenteront alors à nous tout naturellement.

L'ensemble de toutes les fréquences (généralement désignées par ν et exprimées en hertz, ou en cycles par seconde), c'est-à-dire encore celui de toutes les longueurs d'onde (désignées par λ), est donc appelé *spectre*. Il est commode de décrire le spectre électromagnétique au moyen d'un tableau gradué en ν et λ. Au cours du survol de ce tableau, nous ralentirons au-dessus de quelques-unes des cases qui deviennent de plus en plus familières dans le langage courant par leur usage quasi journalier. Il serait hors de propos de tenter une description répertoriée et exhaustive des ondes de toute longueur ou seulement même prétendre à l'étude détaillée d'une région particulière du tableau. Les applications présentées tout au long du premier tome rendent beaucoup mieux compte des possibilités de l'électromagnétisme.

des ondes cérébrales au rayonnement gamma

Il est facile de concevoir un phénomène périodique dont la fréquence serait infiniment basse, un phénomène périodique qui ralentirait à l'extrême, restant un temps extrêmement long dans le même état; on tendrait vers la notion de continu : fréquence nulle, longueur d'onde infinie.

Sans aller jusque-là, on sait détecter et traduire des ondes de très basse fréquence, celles qui sont émises par le cerveau, véritable générateur électromagnétique. La pratique des encéphalogrammes, aussi bien dans le diagnostic des lésions que dans l'examen des troubles psychiques, n'effraie plus le patient, sur qui le médecin branche un nombre impressionnant de « bigoudis-électrodes ». La nature de ces ondes est électromagnétique; leur fréquence, de l'ordre d'une dizaine de cycles par seconde, correspond à des longueurs d'onde de plusieurs dizaines de milliers de kilomètres.

Poursuivons l'exploration des longueurs d'onde élevées. On ne réalise généralement pas quelle est la longueur d'onde qui correspond à notre bon courant alternatif, celui que l'on appelle toujours le « 50 périodes ».

La réponse est facile (en choisissant le kilomètre comme unité de longueur) :

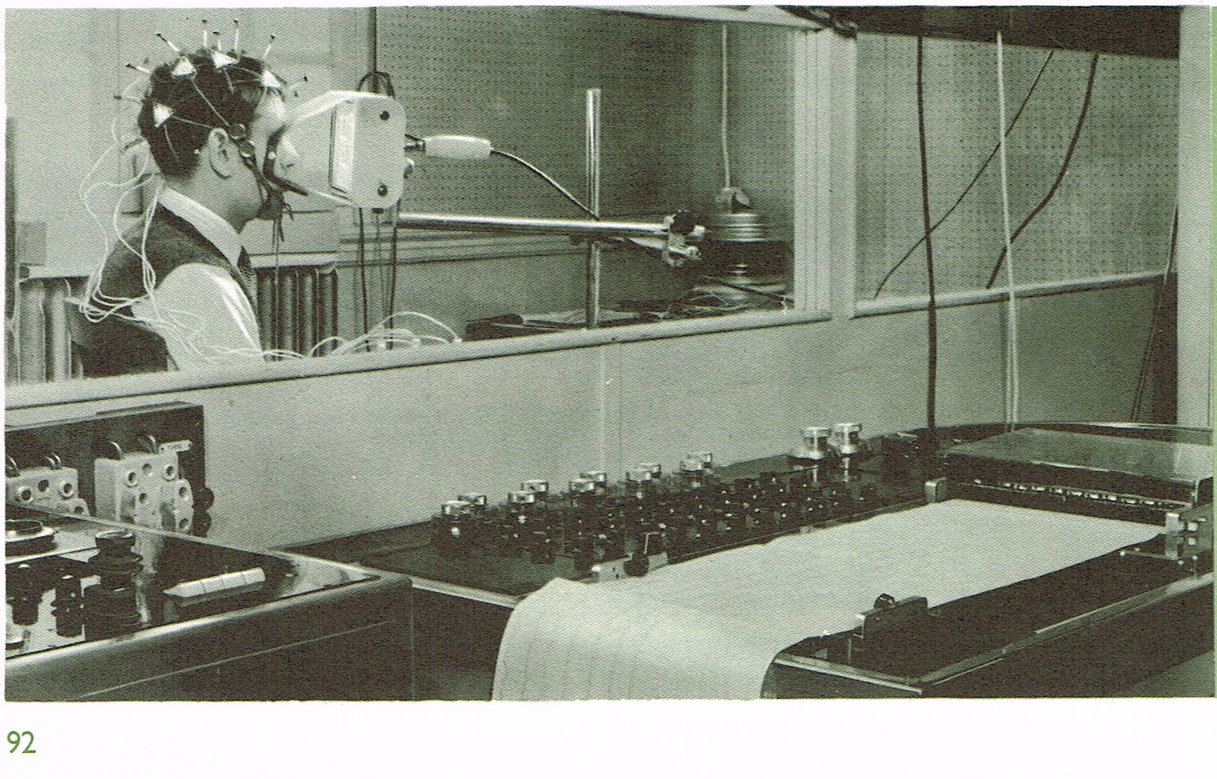
$$\lambda = \frac{c}{T} = \frac{300\,000}{50} = 600 \text{ km.}$$

Signalons aussi un domaine voisin, souvent passé sous silence : celui des longueurs d'onde situées dans la bande des 15 à 20 km. Cette bande est réservée aux communications avec les sous-marins en plongée. La transmission d'ondes électromagnétiques est possible, même à travers une certaine épaisseur d'eau; mais l'absorption, proportionnelle à la racine carrée des longueurs d'onde utilisées, est trop importante si on ne se cantonne pas dans cette bande.

En continuant à descendre dans l'échelle des longueurs d'onde, nous en arrivons aux ondes kilométriques, celles de la radio. Ses pionniers utilisèrent, dès les années qui ont précédé la Première Guerre mondiale, des fréquences comprises entre 15 et 300 kHz, correspondant à des longueurs d'onde de 1 à 20 km (« grandes ondes »). Vite, ils arrivèrent à des ondes plus courtes, de moins de 500 m, dont la propagation avait été reconnue aisée pendant la nuit (la première liaison entre Nice et les Etats-Unis fut réalisée, en 1923, sur ondes d'une centaine de mètres; à la même époque, deux ans plus tard seulement, étaient effectués les premiers essais de télévision, avec une définition de 30 lignes).

A l'autre extrémité du spectre, du côté du rayonnement γ, les longueurs d'onde sont de l'ordre du fermi, soit à peu près de la dimension d'un atome moyen, et les fréquences sont tellement élevées qu'il faut aligner plus de vingt chiffres pour les écrire.

Entre ces extrémités, le spectre ne présente aucune discontinuité. Cependant, les observateurs comme les savants au labo-

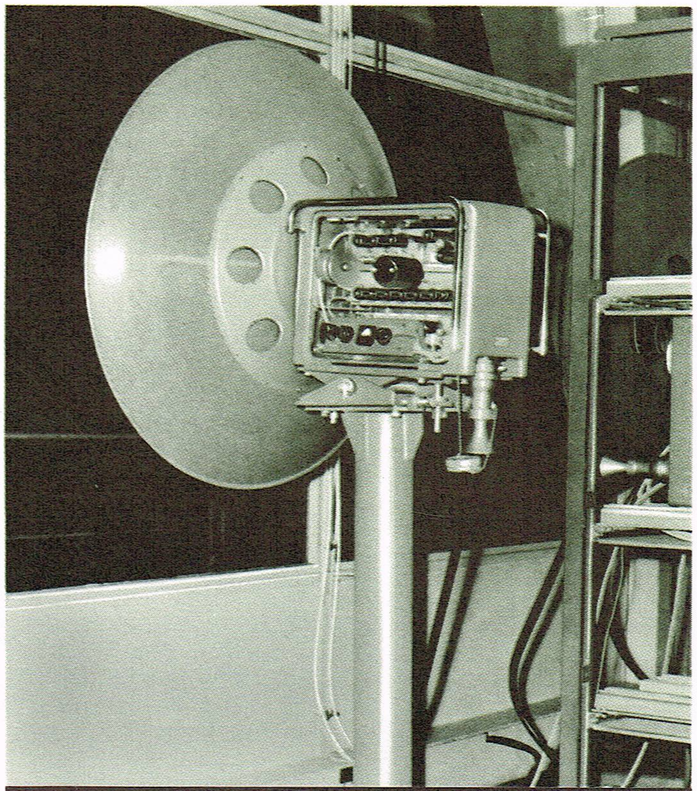


L'encéphalographe enregistre les ondes de basse fréquence émises par le cerveau et renseigne ainsi le praticien sur les lésions ou sur l'origine des troubles psychiques du patient.

Centre phot. de l'A. P.

ratoire n'utilisent qu'un très petit nombre de longueurs d'onde, groupées dans les « fenêtres » auxquelles il a été fait allusion. Donnons-nous une première idée de ces « fenêtres ».

Le rayonnement correspondant à des longueurs d'onde supérieures à quelques dizaines de mètres est arrêté par l'ionosphère. Un autre écran interdit le passage des ondes inférieures à quelques millimètres. L'oxygène, la vapeur d'eau, principalement, absorbent ces radiations. Entre ces deux limites, c'est le domaine de la radio-astronomie (pratiquement celui des λ compris entre 1 mm et 30 m). Ainsi, notre vieille connaissance, le radiotélescope géant de Nançay, dont la surface utile, la plus grande du monde pour des appareils de ce genre, atteint 4 000 m^2, travaille sur 21 cm; cette longueur d'onde, qui est celle d'une raie d'émission de l'hydrogène neutre stellaire, permet l'étude de la structure et de la genèse des galaxies. Un peu plus bas dans l'échelle des longueurs d'onde s'ouvre une nouvelle fenêtre comportant quelques bandes dans l'infrarouge, région du spectre qui s'étend à peu près de 100 μ (infrarouge lointain) à 0,8 μ environ (limite du rouge visible), puis le rayonnement visible, la bande que découpent nos sens, qui va en gros de 0,7 μ à 0,4 μ, et le proche ultraviolet (notons que l'air brumeux est relativement plus transparent pour certaines longueurs d'onde — de 0,8 à 1,4 μ; de 9 à 14 μ et au voisinage de 3,4 μ — que pour la bande du visible; grâce à l'emploi de plaques sensibles particulières, on a pu percer 100 km de brume et obtenir des photos en infrarouge, fort utiles aux militaires). Enfin, au-dessous de 0,29 μ (2 900 Å), aucun rayonnement électromagnétique ne peut pratiquement atteindre le sol; il est alors complètement absorbé par l'ozone de la haute atmosphère (l'ozone est une sorte de condensat chimique de l'oxygène; son odeur caractéristique se reconnaît très bien au voisinage des lignes de transport d'énergie électrique à haute tension).



Relais hertzien.

largeur de bande et information transmise

Il ne semble pas bien nécessaire de s'appesantir sur l'histoire des communications radio-électriques pour réaliser pleinement au moins l'une des raisons qui ont conduit à l'exploitation des longueurs d'onde de plus en plus courtes. (Il est cependant d'autres raisons que celle qui est évoquée ici : par exemple leur moindre affaiblissement lors de la propagation ionosphérique, leur directivité, etc.) Un coup d'œil à un cadran du poste de radio suffit : combien peut-on capter de stations en « grandes ondes »? Tout au plus une dizaine, pour une bande d'à peu près 1 000 m de largeur (de 1 000 à 2 000 m). Et en ondes moyennes? Cinquante, cent peut-être, pour une largeur de 350 m. Regardons enfin la « bande étalée des 49 m » : il semble qu'une foule d'émetteurs se pressent dans un intervalle de longueurs d'onde plus petit que le mètre. Pourquoi cette accumulation? C'est qu'à une valeur donnée de l'écart relatif des fréquences utilisées par deux stations, la largeur de bande effectivement utilisée est d'autant plus faible que la longueur d'onde diminue. Si l'on considère, par exemple — ce qui est généralement admis — que deux ondes dont les fréquences diffèrent d'une dizaine de kilohertz sont susceptibles de transmettre la parole sans « brouillage », on peut fort bien calculer le nombre d'émetteurs pouvant prendre place en ondes longues ou dans la bande des 49 m : λ compris entre 1 000 et 2 000 m correspond à des fréquences ν comprises entre 300 et 150 kHz, soit une largeur de bande de 150 kHz. Ainsi, quinze émetteurs suffisent à encombrer totalement cette gamme. Pour 1,20 m d'étalement, au voisinage de 49 m, la largeur de la bande de fréquences est encore de 150 kHz, et permet donc de « caser » autant d'émetteurs que sur toute la gamme habituelle des ondes longues. Que penser du nombre effarant de communications simultanées qu'il est possible de transmettre en faisceau hertzien (un tube unique serait capable de véhiculer en ondes millimétriques quelque deux mille conversations téléphoniques simultanément)?

Alors, pourquoi ne pas utiliser, d'emblée, une bande très étroite de très petites longueurs d'onde? C'est que les problèmes techniques posés par la production de telles ondes, par leur transmission (nécessité des relais, etc.) et par leur réception sont le plus souvent si ardus qu'à notre époque, pour certains domaines de λ, on en est encore à l'âge des balbutiements. Cela n'est qu'une allusion aux ondes du radar optique, au laser, dont tout le monde parle depuis 1960 et sur lequel tant d'espoirs ont été fondés pour l'avenir des télécommunications; les longueurs d'ondes mises en œuvre étant voisines du micron, ou inférieures, il est facile de calculer, comme tout à l'heure, la bande passante correspondant, par exemple, à la zone du visible (0,8 μ à 0,4 μ) :

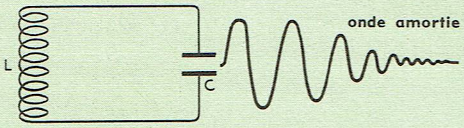
$$\lambda_1 \nu_1 = \lambda_2 \nu_2 = 3.10^{14} \mu; \; \Delta\nu = \frac{3.10^{14}}{0,4} - \frac{3.10^{14}}{0,8} = 3,75.10^{16} \text{ Hz}.$$

Le nombre de messages (et par conséquent d'informations) susceptibles d'être transmis simultanément apparaît effarant. Nous renvoyons, à ce sujet, au chapitre consacré aux lasers, dans le premier tome.

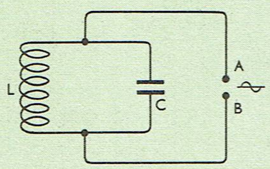
circuits oscillants

Revenons vers le haut de notre échelle des longueurs d'onde, un instant abandonnée, et essayons de nous représenter, le plus schématiquement possible, la fabrication des ondes du type radio. Il a souvent été question jusqu'ici de secouer des électrons, de détacher des ondes de l'antenne comme des ronds de fumée; l'élément de base est le circuit oscillant. Un message étant traduit en variations de courants ou de tensions, ces variations entraînent la génération d'ondes électromagnétiques.

Service Documentation-Publicité.

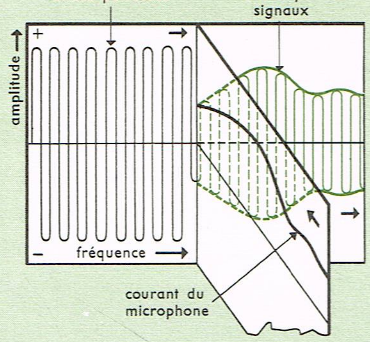


self et capacité : deux circuits élémentaires (ondes amorties)

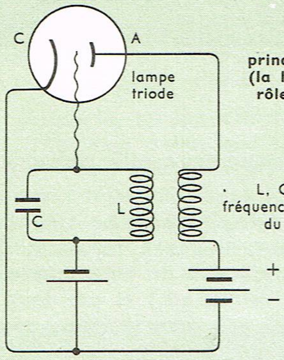


circuit oscillant (ondes entretenues)

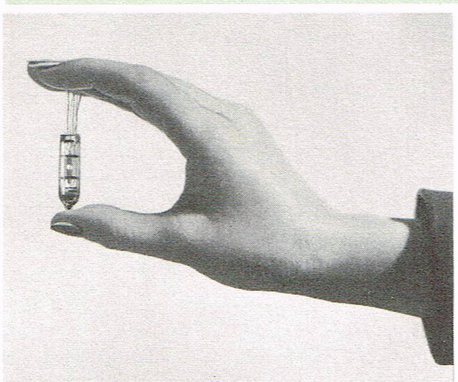
principe de la modulation des ondes électromagnétiques



principe de l'hétérodyne (la lampe triode joue le rôle d'auto-oscillateur)



Comment les ondes radio-électriques sont produites et modulées.





Un tube à vide du type miniature et un klystron.

En quoi consiste donc ce circuit oscillant? Le plus simple imaginable est représenté ci-contre. Il comporte un condensateur de capacité C et une self-inductance L dont la résistance ohmique est le plus faible possible. Supposons le condensateur initialement chargé; il se déchargera dans la self qui, on le sait, présente une inertie se manifestant par la création d'un courant s'opposant au courant inducteur — c'est-à-dire au flux d'électrons déversés lors de la décharge de la capacité, et qui, au passage dans la self, engendrent un champ magnétique d'induction. Le courant de self tendra donc, avec un certain retard, à recharger le condensateur, et le cycle se poursuivra : c'est la décharge oscillante classique. Cependant, par suite des pertes dans le circuit, dont la plus importante est due au fait que la résistance ohmique de la self n'est pas rigoureusement nulle, le mouvement oscillatoire s'amortirait s'il n'était « entretenu » par l'apport énergétique d'une source alternative extérieure, connectée en AB. Pour des valeurs de L et C données, les conditions optimales d'entretien sont réalisées pour une fréquence particulière, la fréquence de résonance. A cette fréquence, l'impédance du circuit est minimale (1). Lorsque le circuit placé sur l'antenne a été accordé (par exemple, en réglant la capacité) sur une fréquence désirée et se trouve ainsi dans les conditions de résonance, la fréquence de l'onde électromagnétique rayonnée par l'antenne est essentiellement celle de résonance.

Mais ne perdons pas de vue le pourquoi de ce système oscillant : nous voulions produire, transmettre ou recevoir des ondes radio. Nous étions même axés sur la radiodiffusion. La parole, les sons qui doivent être reproduits correspondent à des fréquences trop basses (les audiofréquences) pour être transmis directement. Aussi, lorsqu'on énonce que « la transmission est effectuée sur une longueur d'onde de 300 m », cela veut dire qu'après avoir traduit les vibrations mécaniques (son ou parole) en signaux électriques de basse fréquence, on a véhiculé ceux-ci de l'émetteur au récepteur au moyen d'une « onde porteuse » de 300 m de longueur, « modulée » par l'onde de basse fréquence.

Le véhicule, l'onde porteuse, est créé sur place par un oscillateur comprenant, outre des circuits du type précédent, un élément nouveau : les lampes électroniques (de plus en plus remplacées par les transistors pour des raisons que nous avons vues longuement dans le premier tome), encore appelées « tubes à vide » parce qu'elles sont constituées d'une ampoule à l'intérieur de laquelle règne un vide poussé et où sont installées des électrodes aux fonctions particulières (cathodes émettrice, grille de commande, anode collectrice, etc.). La technologie des circuits électroniques ne faisant pas l'objet de ce chapitre, disons seulement que ce genre de lampe est couramment utilisé pour des radiofréquences allant jusque vers les ondes métriques, bien qu'il présente de sérieuses limitations d'emploi dans le domaine des ondes ultra-courtes. En fait, au-dessous des longueurs d'onde de l'ordre d'une dizaine de mètres (fréquence de l'ordre de 30 millions de hertz), les tubes oscillateurs cèdent le pas à des tubes à cavité résonnante ou à modulation de vitesse, dans lesquels les électrons sont groupés en paquets très denses, à l'encontre de ce qui se passe dans la triode classique (2). Les énergies haute fréquence mises en jeu peuvent ainsi être portées à des valeurs considérables.

Avec ce type d'oscillateurs, les klystrons par exemple, nous nous introduisons dans la région des ondes décimétriques et centimétriques, dont les générateurs font intervenir, le plus souvent, à la fois le champ électrique et le champ magnétique (magnétrons, tubes à ondes progressives). C'est le domaine des hyper-

(1) Les impédances de la self et de la capacité étant respectivement proportionnelle et inversement proportionnelle à la fréquence, on peut s'attendre que les effets de C et L s'équilibrent pour une fréquence déterminée (résonance).

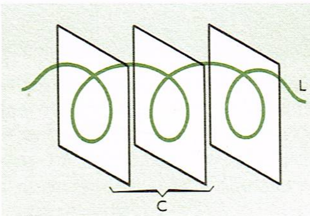
(2) Le temps de transit des électrons entre les électrodes d'une telle lampe étant du même ordre de grandeur que la période d'oscillation, on conçoit combien un regroupement des charges électriques s'avère nécessaire.

Phot. Compagnie générale de T. S. F. et Thomson-Houston.

fréquences, de la technique des guides d'ondes et des cornets, des types de propagation qui se rapprochent de celui de la lumière, de cette application, enfin, qui ne semble même plus étonner aujourd'hui : le radar.

le radar, vers des λ de plus en plus petits

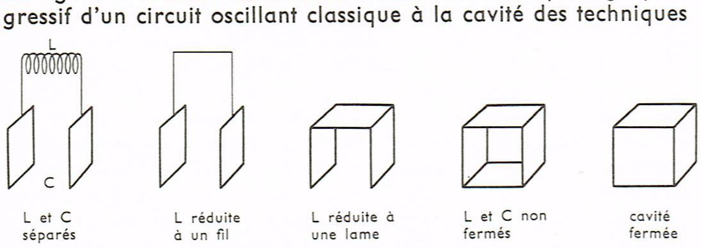
Nous avons maintenant fait un bond énorme depuis les fréquences de la radiodiffusion : nous parlions en kilohertz, nous énoncerons désormais les fréquences en mégahertz (MHz), c'est-à-dire en millions de cycles par seconde. Déjà les fréquences des signaux utilisés pour la transmission des images de la télévision étaient comprises entre 30 et 300 MHz, correspondant à des longueurs d'onde de 1 à 10 m (bande de 185,25 MHz pour la « modulation vidéo »). Quant au radar (abréviation de « RAdio Detection And Ranging »), ses « tops » ont des durées de l'ordre du millionième de seconde; il s'agit de trains d'ondes de fréquences de plusieurs milliers de MHz. A ces fréquences, les circuits oscillants ne sauraient se concevoir d'une façon



Deux boucles voisines se comportent comme une capacité très petite, surtout s'il s'agit d'ondes de radar.

aussi élémentaire que ceux de la radio, puisqu'une self extrêmement faible possède alors une grande impédance, tandis que la capacité reste petite. En d'autres termes une petite boucle de conducteur réduite se comportera comme une self à forte impédance, et le courant passera facilement d'une boucle à la suivante par un effet capacitif.

La figure ci-dessous schématise excellemment le passage progressif d'un circuit oscillant classique à la cavité des techniques

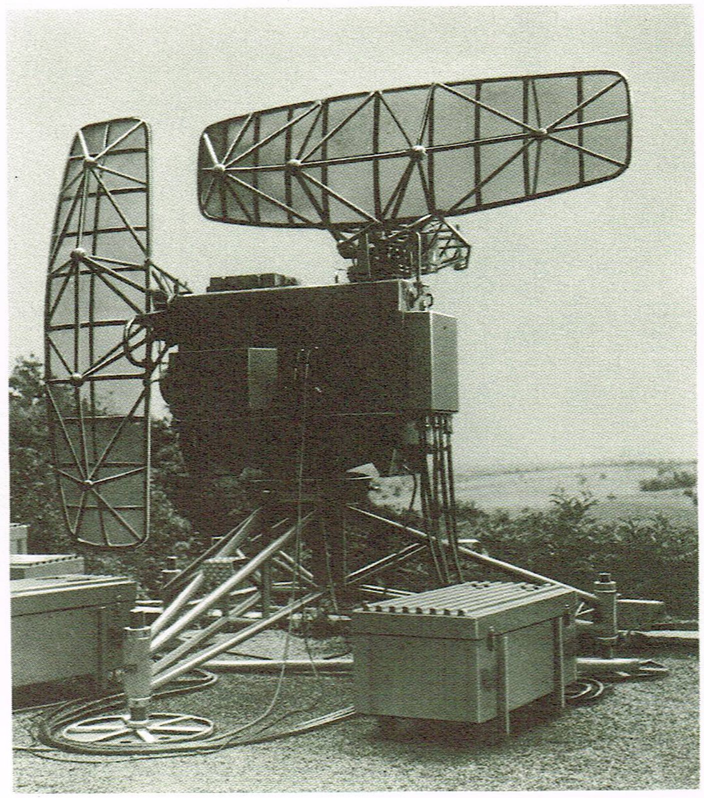


Passage du circuit oscillant classique à la cavité (d'après Champeix).

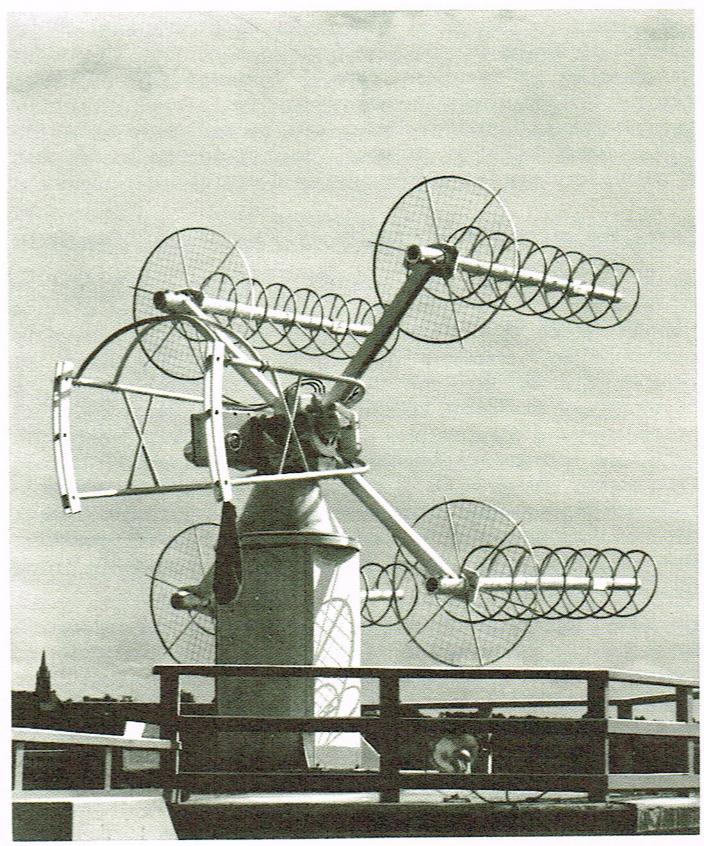
centimétriques. Le transport même de l'énergie électromagnétique sera réalisé au moyen de conducteurs creux, les **guides d'ondes,** dont le domaine d'utilisation s'étend jusqu'aux longueurs d'ondes de l'ordre du millimètre. Dès lors, nous arrivons à la limite de l'optique de l'infrarouge, sinon même du visible.

En descendant de plus en plus l'échelle des longueurs d'onde, nous verrons évoluer le langage et la technologie. Sans empiéter sur ce que nous verrons un peu plus loin, remarquons dès maintenant que la liaison effective a déjà été réalisée il y a quarante ans, lorsque des ondes hertziennes de $\lambda = 0,22$ mm furent obtenues par Nichols et Tears, tandis que, de son côté, l'optique permettait d'observer également des longueurs d'onde de même ordre, sinon supérieures, dans l'infrarouge lointain ($\lambda = 0,313$ mm à partir d'une lampe à mercure).

Nous avons mentionné le radar comme une grande réalisation des hyperfréquences. On met habituellement en relief l'analogie de son principe, d'une part avec l'écho et de l'autre avec le système de guidage de la chauve-souris, dont les signaux

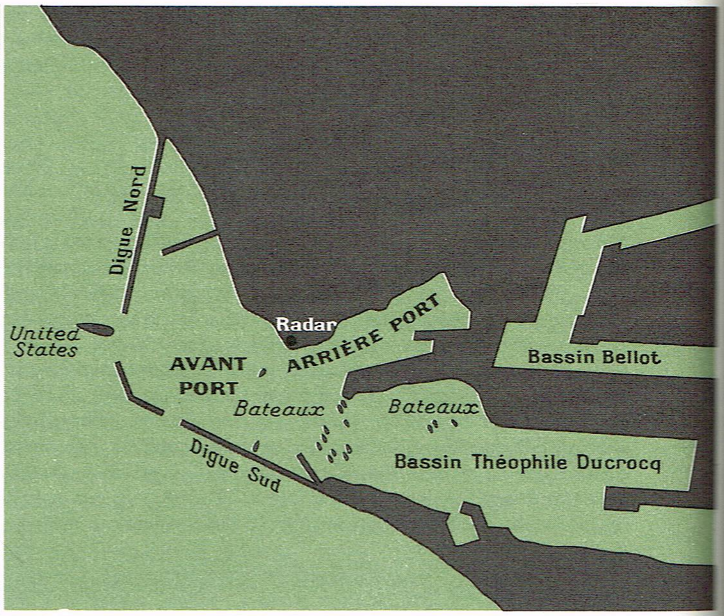


Radar léger d'aide à l'atterrissage des avions et, *ci-dessous,* radar aérien complexe pour assurer les télécommunications avec les satellites-relais.



Phot. C. F. T. H. et R. T. F.





La zone portuaire du Havre photographiée sur l'écran du radar qui équipe ce port, et dont l'emplacement est signalé par un point vert au centre du plan. A gauche, le navire *United States* quittant le port.

ultra-sonores vont se réfléchir sur l'obstacle et sont captés au retour par l' « émetteur-récepteur » dont use ce petit animal pour se diriger dans l'obscurité. Cette image, pour approximative qu'elle soit, n'en est pas moins correcte. L'émetteur radar d'un navire balaie les différents azimuts de signaux électromagnétiques (et non plus d'ondes classiques comme pour la chauve-souris) émis sous la forme de traits pendant des intervalles de temps suivis de silences. Ces signaux périodiques sont choisis de fréquence élevée afin de mettre à profit les qualités de directivité attachées à ce rayonnement. L'écho sur les obstacles qui surgissent dans le voisinage d'un navire — que celui-ci soit équipé pour la chasse de la baleine ou pour la protection antiaérienne des convois, ou qu'il tente d'emprunter un chenal à l'entrée du port — permet alors de les détecter, et de les localiser (en direction et en distance).

fenêtre atmosphérique et radio-astronomie

Avant de quitter les hyperfréquences, n'est-il pas tentant de soulever un coin du voile derrière lequel se cachent d'autres mondes, et de se demander quelles informations elles nous apportent en véhiculant des signaux extra-terrestres? Les radioastronomes étudient aujourd'hui systématiquement le rayonnement des corps célestes, dont les premières émissions hertziennes ont été détectées il y a un peu plus de trente ans. Ainsi, on sait maintenant observer une raie d'émission de l'hydrogène stellaire, de 21 cm de longueur d'onde (toute la radioastronomie, que nous avons traitée longuement à deux reprises dans deux chapitres du tome I, repose sur la constatation que l'atmosphère laisse passer cette radiation, de grand intérêt cosmique).

De quoi s'agit-il? Les deux particules élémentaires qui constituent l'atome d'hydrogène — l'électron et le proton — se comportent un peu comme des aimants qui passeraient d'une position parallèle à une position antiparallèle, c'est-à-dire de sens différent. Ce changement de position se traduit par un changement de situation énergétique (qu'on appelle « transition »), auquel correspond une émission : celle du rayonnement électromagnétique de longueur d'onde 21 cm dans le cas présent. Or, les théoriciens savent démontrer que cette transition spontanée de l'atome d'hydrogène se produit en moyenne tous les 11 millions d'années. Si donc l'observateur est à même de repérer l'émission de la raie 21 cm et son intensité relative en différents points du ciel, il pourra tracer une carte sur laquelle il portera point par point les densités d'hydrogène déduites des intensités. Les mouvements de chaque région observée sont déterminés par l'étalement des fréquences relevées; en effet, l'émission n'apparaît pas limitée à une fréquence unique, et la raie n'est pas infiniment étroite, mais se trouve élargie dans le spectre par effet Doppler-Fizeau, dont les manifestations sont d'autant plus marquées que les déplacements relatifs de la source et de l'observateur sont plus rapides. Des nébuleuses, des masses gazeuses qui se promènent dans l'univers à des vitesses atteignant quelques milliers de kilomètres par seconde quoique invisibles optiquement restent accessibles à nos moyens humains. Le « profil » de la raie représente donc finalement, compte tenu de l'effet Doppler-Fizeau, l'intensité du rayonnement émis en fonction de la vitesse de rapprochement ou d'éloignement de l'émetteur. C'est de cette manière qu'on a réussi à dresser, au cours de ces dernières années, une véritable carte de notre Galaxie — de cet ensemble de myriades d'étoiles dont notre Soleil n'est qu'un individu —, et à nous représenter l'immense spirale aux bras multiples dans laquelle notre monde est entraîné.

Phot. U. S. I. S.

DE L'INFRAROUGE A L'ULTRAVIOLET

Il n'est pas jusqu'à la cuisine moderne qui ne se laisse envahir par la technologie. On a beaucoup parlé du chauffage électronique, de la cuisinière à magnétron. N'oublions pas que l'onde électromagnétique reste un transporteur d'énergie et que la puissance mise en jeu est d'autant plus considérable que la fréquence est grande; au lieu d'être dirigée vers une antenne qui la rayonnera au loin, l'énergie est ici transformée en chaleur à l'intérieur de la masse à échauffer (chauffage par pertes dans les diélectriques). Le quart de poulet qui vient d'attirer l'œil et l'estomac affamé du dîneur attardé a été paré longtemps à l'avance. Notre client le considère tout à loisir à travers la vitre du compartiment frigorifique où il pourrait attendre des jours encore. Qu'il le choisisse : en quelques dizaines de secondes, le réchauffeur hyperfréquence — 500 ou 600 millions de cycles par seconde — permettra de lui servir l'objet de sa convoitise, à la peau dorée à souhait.

Chaque jour voit s'étendre la gamme d'utilisation du chauffage obtenu par la technique des hyperfréquences.

A la diathermie, technique d'échauffement des tissus vasculaires, et qui utilise des longueurs d'onde d'une centaine de mètres, se substitue de plus en plus une thérapeutique mettant en œuvre les hyperfréquences (100 à 1 000 MHz environ), dont le champ d'application apparaît extrêmement vaste, allant du traitement des rhumatismes à la stimulation des glandes endocrines, de l'apaisement des douleurs névralgiques aux interventions gynécologiques.

Mais d'autres applications des hyperfréquences sont possibles; imagine-t-on par exemple que ces ravissants petits bateaux aux coques multicolores qui évoluent si légers dans maint rêve de vacances ont été moulés et leurs éléments soudés avec des ondes électromagnétiques de 200 MHz?

Mais le lecteur ne va-t-il pas se lasser d'une interminable énumération? Ne va-t-il pas s'essouffler, perdu dans l'écheveau de toutes ces ondes qui s'entrecroisent de quelque côté qu'il se tourne? Parce que l'aventure n'est pas terminée. Nous avons à descendre de plusieurs échelons encore vers la cascade des λ de plus en plus petits.

l'infrarouge (de 1 000 à 1 micron)

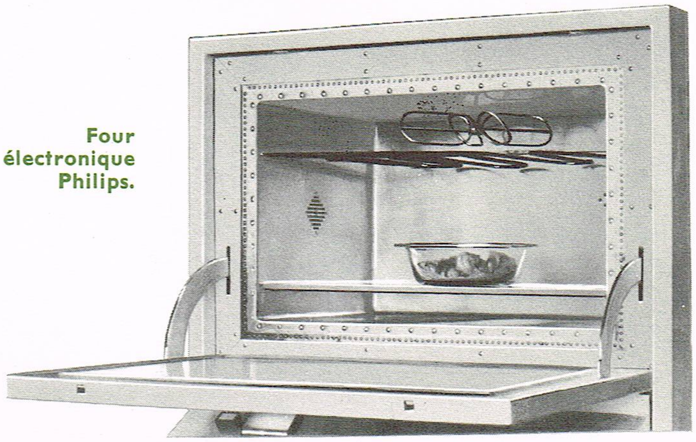
L'infrarouge commence dans la région des ondes millimétriques et va jusqu'au visible. C'est un peu par une exagération du langage qu'on l'a baptisé « rayonnement calorifique ». Sans doute produit-il de la chaleur, permet-il un dégagement d'énergie calorifique, est-il souvent détecté par des effets thermiques. Mais, il en est ainsi de beaucoup d'autres rayonnements : lorsqu'on amène sur un morceau de papier l'image du soleil, bien visible, au moyen d'une grosse loupe, le papier noircit vite ou même prend feu; lorsqu'on place la main devant un cornet ultra-hertzien ou dans un faisceau radar intense, la sensation de chaleur n'est pas non plus négligeable. Le rayonnement infrarouge agit sur la structure même des molécules, sur l'édifice qui constitue la matière, dont les éléments s'agitent, se bousculent, entrent en mouvement de rotation ou de vibration sans qu'il soit besoin de conduction ou de convection thermiques.

L'énergie électromagnétique est donc simplement absorbée, et il y a un échauffement dans la masse.

rayonnement et température

C'est ce qui se passe pour la plupart des corps, qui présentent presque toujours une absorption forte aux longueurs d'onde de l'infrarouge. Réciproquement d'ailleurs, la matière chauffée

Four électronique Philips.



émet un rayonnement électromagnétique dont les caractéristiques varient, pour une substance donnée, avec la température atteinte : l'agitation des molécules, produite au moyen de l'énergie thermique fournie, se traduit par une émission électromagnétique. Tous les corps émettent de l'infrarouge, et non pas seulement les substances portées à l'incandescence. Le corps humain lui-même est émetteur, avec un maximum vers 9-10 μ, sans que nous soyons pour autant très lumineux.

La longueur d'onde de cette émission diminue lorsque la température augmente. La nébuleuse du Crabe, dont la température serait de l'ordre de 500 000 ^{0}K, rayonne principalement dans le violet et l'ultraviolet. Chacun de nous a constaté le changement de couleur du poêle chauffé, qui, du rouge sombre, devient de plus en plus clair et passe au « blanc ». Mais ici, nous sommes déjà dans la région visible du spectre, ou, tout au moins, à sa frontière. Il est difficile, en effet, de marquer la limite inférieure de l'infrarouge en se fondant sur la sensibilité de l'œil : si la température du poêle n'était que de 700-800 ^{0}C, le rayonnement correspondant, de longueur d'onde voisine du micron, ne serait alors perçu que si la quantité d'énergie rayonnée est très grande. On éprouve une sensation de rouge, due au fait que la rétine réagit devant le flux d'énergie au-delà de ses possibilités habituelles.

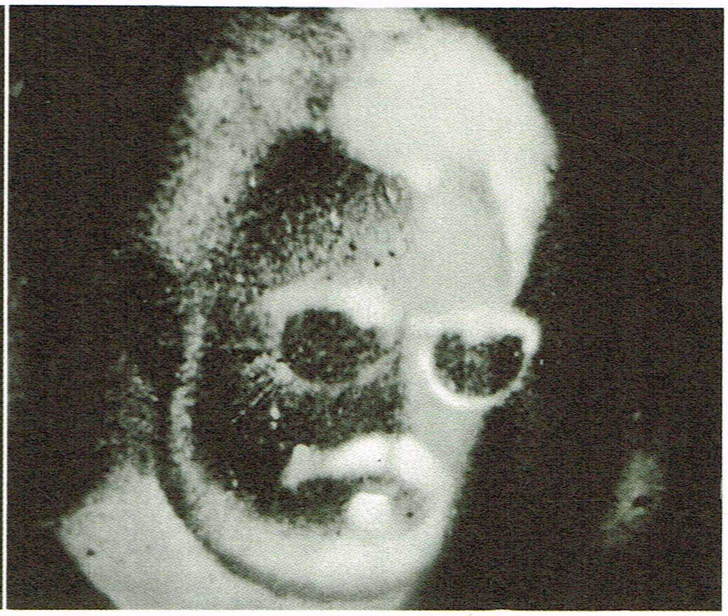
Cet infrarouge, lié si étroitement à ses effets de chaleur, est entré dans l'histoire vers 1800, en se manifestant à l'astronome anglais Herschel. Explorant le rayonnement du Soleil au moyen de verres différemment colorés qu'il plaçait sur le faisceau de lumière, il s'aperçut que l'impression variait grandement avec la couleur. Lorsqu'il promenait un thermomètre tout au long du spectre, les indications étaient différentes d'une extrémité à l'autre et l'instrument réagissait bien au-delà de la région « lumineuse ».

Comme tout autre domaine du spectre, l'infrarouge transporte de l'énergie transformable par la matière inerte. La température indiquée par le thermomètre de Herschel s'élève alors même qu'on a quitté la gamme des infrarouges.

On peut isoler une bande en se servant de filtres, en l'occurrence de verres colorés, qui ne laissent passer que certaines longueurs d'onde et absorbent les autres. Ainsi, les verres deviennent-ils opaques vers 2 à 3 μ, le quartz à 4,5 μ, la fluorine vers 9 μ, le sel gemme (sans impuretés) vers 15 μ. Au-delà d'une vingtaine de microns, l'air lui-même perd sa transparence, qu'il reprend du reste dans les longueurs d'ondes hertziennes. Les outils utilisables en infrarouge vont donc se différencier suivant la région mise en cause. Les détecteurs devront présenter soit une grande souplesse, soit une subtile diversité. Les applications n'en seront que plus riches!

Phot. Philips





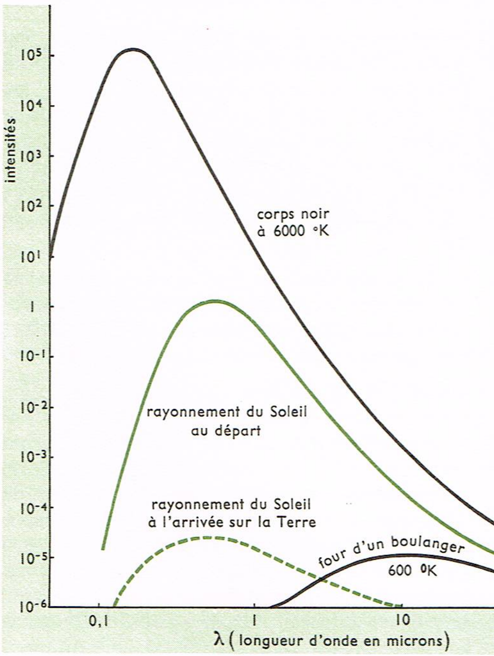
Deux portraits d'un même personnage; *à gauche,* en lumière visible; *à droite,* en infrarouge par photographie d'un écran de télévision, donnant l'image prise par un tube analyseur sensible seulement aux rayons infrarouges de 1 à 1,8 micron. Les parties les plus froides du sujet (système pileux, vêtements, lunettes) donnent une teinte claire, alors que la peau apparaît en noir.

Pour entrer dans ce concret, considérons d'abord les courbes ci-dessous. Suivant l'axe des abscisses, on a porté les longueurs d'onde en échelle logarithmique, c'est-à-dire que l'intervalle représentant la différence entre deux valeurs de λ reste le même lorsqu'on passe d'un multiple de 10 à l'autre (1). Les intensités sont exprimées, à partir d'une unité arbitraire, en puissances de 10 suivant l'axe des ordonnées. Les températures sont données en degrés Kelvin. La courbe supérieure représente le rayonnement d'un corps idéal, théorique (corps noir), qui serait porté à 6 000 °K; la seconde est celle du rayonnement solaire correspondant à la même température au moment de l'émission. La courbe pointillée montre ce qu'il en reste après les pérégrinations du voyage Soleil-Terre, les absorptions diverses, en un mot l'amortissement. Enfin, l'autre courbe, nettement décalée, serait celle de l'émission de notre poêle de tout à l'heure, lorsque sa température ne dépasse guère 300 °C (c'est à peu près l'ambiance à l'intérieur du four du boulanger, lorsque l'on y prépare des flûtes croustillantes), soit sensiblement 600 °K.

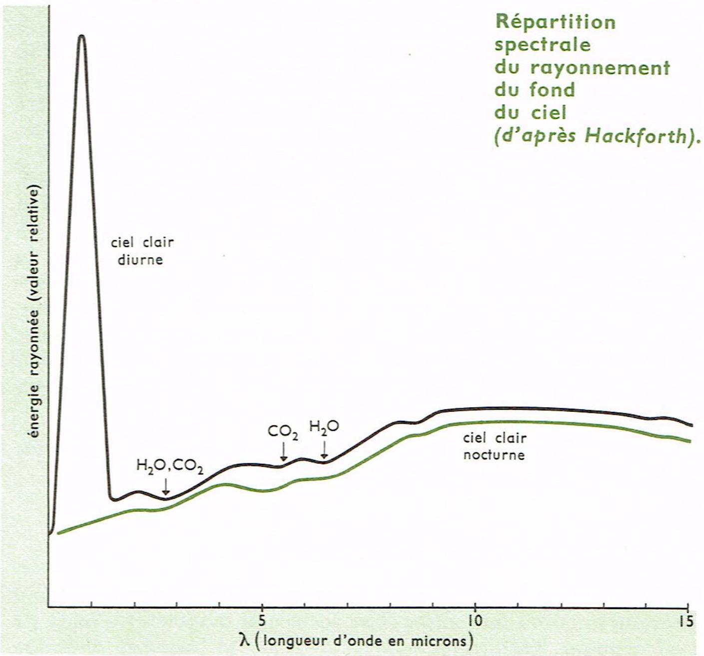
Il vient d'être fait mention du corps noir (1). Il sera étudié plus en détail dans une autre partie du présent ouvrage. Indiquons seulement qu'il existe un rapport parfaitement défini entre l'énergie émise et l'énergie absorbée par un corps à température donnée (loi de Kirchhoff), et cela quelle que soit la longueur

(1) Rappelons que log 10 = 1, log 100 = log 10^2 = 2 log 10 = 2, log 10^3 = 3 log 10 = 3, etc.; de même, log 0,1 = log 10^{-1} = — log 10 = — 1, log 0,01 = log 10^{-2} = — 2 log 10 = — 2, etc.; d'où la graduation commode en puissances de 10.

(1) Une sphère creuse dont l'intérieur serait recouvert de noir de fumée, et percée d'une ouverture réduite, est un « corps noir ». L'intérieur d'un four-creuset se comporte à peu près comme le corps noir théorique.

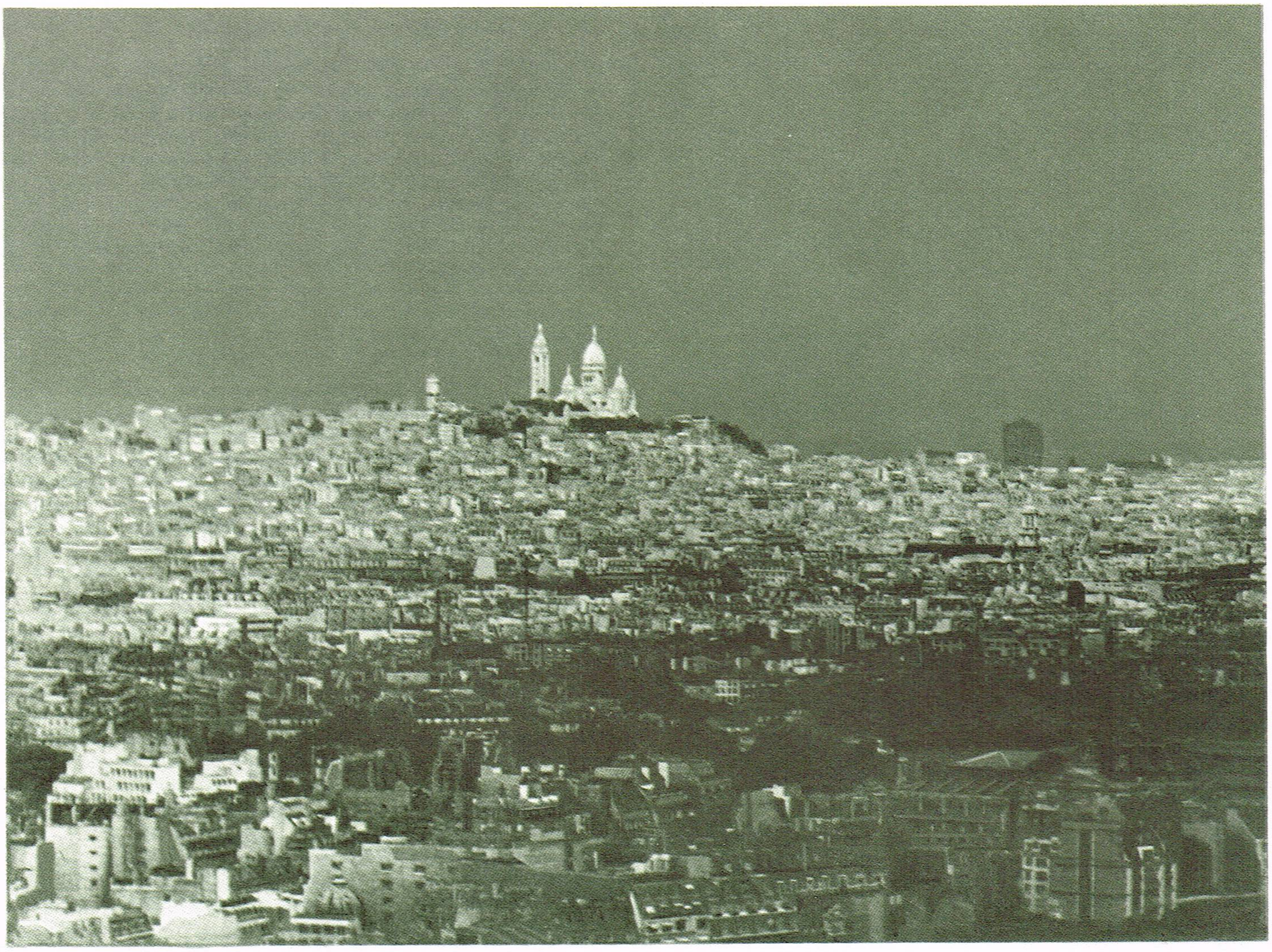


Distribution d'énergie du rayonnement thermique en fonction de λ aux différentes températures *(d'après Gamow).*



Répartition spectrale du rayonnement du fond du ciel *(d'après Hackforth).*

Phot. L. E. P.



Le Sacré-Cœur pris avec un téléobjectif à longue distance focale muni d'un filtre rouge, depuis le Mont-Valérien.

d'onde. On peut donc s'attendre qu'une substance fortement absorbante rayonne largement : c'est bien ce qui, instinctivement, nous fait choisir en été des vêtements clairs, de préférence aux sombres, sous lesquels « on a plus chaud ». Le corps noir représente l'idéal en manière d'absorption; son émission est donc également maximale : il est capable de réémettre exactement tout ce qu'il a reçu. L'énergie qu'il rayonne pour l'ensemble des longueurs d'onde susceptibles d'émission est proportionnelle à l'aire comprise entre l'axe des λ et la courbe d'intensité à la température T. Cette énergie est proportionnelle à la quatrième puissance de T **(loi de Stefan)**.

On éprouve une telle crainte à l'abord du si vaste champ d'applications des infrarouges, qu'il paraît nécessaire de se faire une idée de quelques types de récepteurs et de leurs variétés suivant la gamme à l'intérieur de laquelle on travaille.

quelques autres applications de l'infrarouge

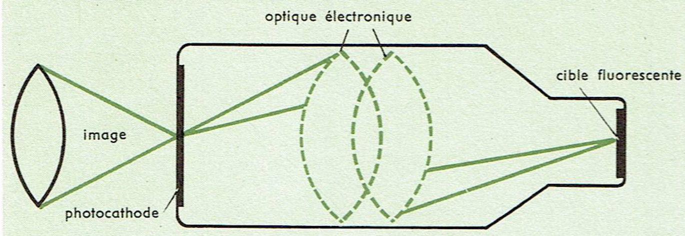
L'infrarouge, on l'a vu, s'étend des longueurs d'onde millimétriques (1 000 μ) au visible (0,8 μ). Cependant, jusqu'à ces dernières années, on ne se servait pratiquement que de la bande des 15-20 μ.

L'infrarouge proche (0,8 — 1,5 μ) ne procède pas de méthodes très différentes de celles du visible. Les cellules photo-émissives (photocathodes de césium sur argent oxydé, principalement dans les photomultiplicateurs courants), les cellules photo-résistantes (à l'oxysulfure de thallium, par exemple), la plaque photographique restent les détecteurs habituellement employés.

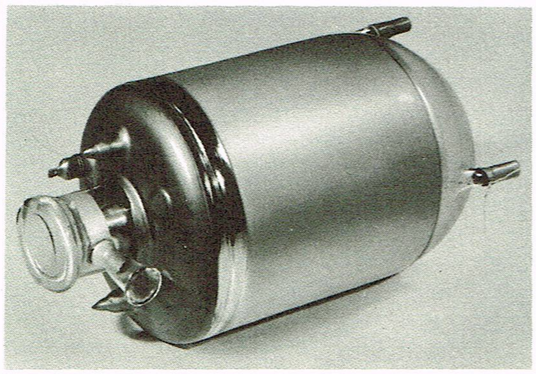
On trouve dans le commerce des émulsions photographiques valables pour cette région du spectre. La faible diffusion de l'atmosphère (proportionnelle à l'inverse de λ^4) permet de prendre des photographies de lointains assez surprenantes, comme on l'a vu déjà. Il ne faut pourtant pas se leurrer : les installations donnant des images à 100 km mettent en œuvre des téléobjectifs à très longues distances focales, qui ne sont guère à la portée de l'amateur.

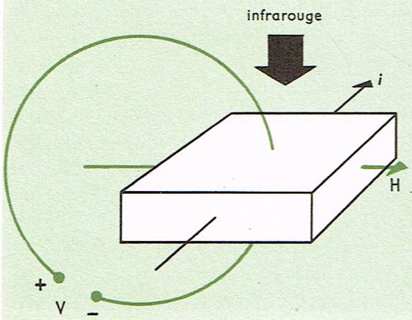
Néanmoins, sans appareillages spéciaux, on a utilisé dans certaines classes pilotes la photographie infrarouge en psychologie appliquée, afin de fixer les réactions des enfants dans l'obscurité au cours de projections d'images ou de films.

Phot. François Merlet, Atlas Photo.



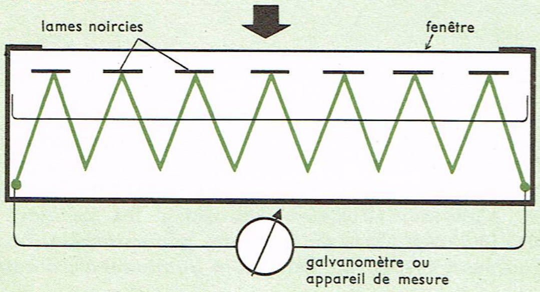
Principe du tube transformateur d'image. ***Ci-dessous,*** **tube transformateur d'image utilisé pour rendre visibles les « images » fournies par des rayons X.**



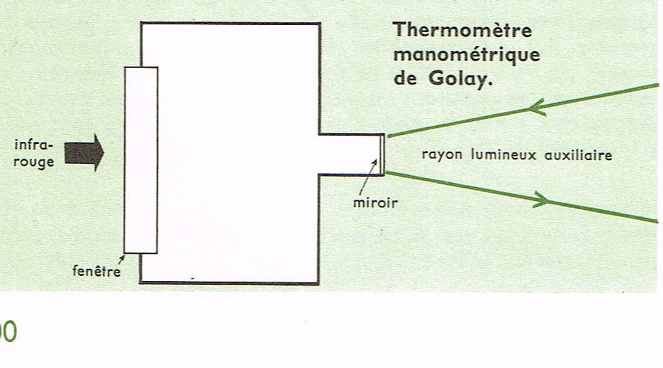


Principe de fonctionnement d'un appareil photo-électromagnétique pour la mesure de l'infrarouge (mise en œuvre de l'effet Hall).

Schéma de principe d'un bolomètre.



Thermomètre manométrique de Golay.



les « télescopes électroniques »

Le convertisseur d'images — technique plus jeune encore due au physicien et astronome A. Lallemand — est bien connu maintenant sous le vocable de « télescope électronique ». C'est un peu cela, après tout, puisqu'on l'utilise en astronomie, puisque des électrons sont expulsés d'une photocathode par le rayonnement infrarouge à détecter. Ces électrons, accélérés au moyen de champs électriques produits par des lentilles électrostatiques, sont triés, regroupés sur un écran fluorescent et restituent l'image — visible cette fois — de l'objet ou de la source infrarouge.

autres détecteurs IR

Ce qu'on appelle « infrarouge moyen » correspond aux longueurs d'onde comprises entre 1,5 et 6 μ. Les photodétecteurs, cellules photorésistantes, au sulfure de plomb par exemple, furent longtemps l'outil de choix. La résistance de ces cellules est fonction de l'éclairement reçu (pour λ déterminé). La sensibilité augmente lorsqu'on travaille à basse température (neige carbonique, azote liquide); la limite d'utilisation s'en trouve généralement repoussée vers des valeurs supérieures des longueurs d'onde.

Actuellement, d'autres détecteurs, très stables et dont la « réponse » reste la même jusque vers 6 μ, sont réalisés à partir de semi-conducteurs (germanium) ou encore en mettant à profit des phénomènes particuliers, tels que l'effet Hall : un système photo-électromagnétique, par exemple, est constitué d'un semi-conducteur (antimoine-indium) placé dans un champ magnétique intense et parcouru par un courant perpendiculaire à ce champ magnétique. L'action du rayonnement infrarouge sur la plaque semi-conductrice fait apparaître une différence de potentiel entre les faces supérieure et inférieure.

Dans l'infrarouge lointain, les récepteurs thermiques sont pratiquement les seuls valables. Les piles thermo-électriques, association de plusieurs couples thermo-électriques placés en série (argent-bismuth, argent-palladium, par exemple), indiquent l'intensité du rayonnement infrarouge incident par la variation de résistance de la chaîne; l'infrarouge produisant une élévation de température, celle-ci affecte directement la résistance de jonction. Soulignons ici une particularité intéressante des récepteurs thermiques : ce sont les seuls qui soient utilisables dans toute l'étendue du spectre électromagnétique, des rayons gamma aux ondes kilométriques.

Les bolomètres sont des fils ou des lames minces de métal noirci (platine, nickel, or), dont la résistance, fonction de l'échauffement (c'est-à-dire de l'énergie infrarouge reçue), est mesurée directement au moyen d'un appareil approprié (pont de Wheatstone).

Les thermistors fonctionnent sur un principe voisin : une lame mince de semi-conducteur dont la résistance présente un coefficient de température élevé est placée dans un circuit dont les variations de résistance se traduisent par des variations de courant, amplifiées convenablement.

Le système de Golay est un thermomètre manométrique : la cloison élastique d'une capsule contenant un gaz (xénon) joue le rôle d'un miroir qui se déforme lorsque les infrarouges dirigés sur la fenêtre de l'instrument provoquent un échauffement du gaz, donc une variation de sa pression. Un pinceau lumineux auxiliaire, réfléchi par la membrane-miroir, « suit » les déformations de celle-ci et rend compte du rayonnement infrarouge.

D'autres moyens de détection et de mesure (évaporographie, phosphores sensibles, convertisseur IR lointain) ne sauraient trouver place dans ce survol déjà si pesant. Quelques applications termineront maintenant notre longue escale.

Doc. Philips.

comment y voir la nuit ?

Il paraît que les buses, ces grands oiseaux qui planent des heures durant au-dessus des campagnes dans l'attente de pouvoir fondre sur le malheureux poussin égaré, doivent leur vue si perçante à leur sensibilité dans l'infrarouge. Ainsi, plus de camouflage possible pour l'oisillon gris ou brunâtre qui se cache sous les basses feuilles. Sa température, plus élevée que celle de son environnement, fait de lui un radiateur d'infrarouge que le rapace aura tôt fait de repérer. Il ne saurait devenir invisible qu'en supprimant tout dégagement de chaleur, c'est-à-dire qu'en se conservant à une température rigoureusement identique à celle du milieu où il évolue. Ne savait-on pas, de même, dès la fin de la Seconde Guerre mondiale, détecter la nuit une cigarette à plusieurs kilomètres, l'échappement d'un avion ou la masse mouvante d'un navire à des dizaines de kilomètres?

Que son emploi soit pacifique ou guerrier, la « visualisation de l'invisible », la vision de nuit (ou par temps de brume, bien que la densité, les dimensions des gouttelettes, les accidents du terrain soient autant de facteurs susceptibles de réduire le pouvoir des infrarouges) peut se concevoir de deux façons : ou bien un faisceau de longueur d'onde déterminée est dirigé sur l'objectif, que l'on examine alors à travers une lunette-convertisseur d'image réglée sur la même bande de longueur d'onde (c'est le cas des phares de véhicules munis de filtres appropriés, le conducteur étant alors équipé de lunettes correspondantes), ou bien l'on compte sur le rayonnement propre de l'objectif, reçu sur un détecteur à large bande.

identification des substances à l'aide des infrarouges

Au début de cette présentation de l'infrarouge, l'accent a été mis sur son interaction avec les éléments, avec la structure même de la matière. Une longueur d'onde déterminée se comporte d'une façon spécifique à l'égard d'une substance donnée. Chaque substance se trouve donc dotée d'une carte d'identité particulière (effectivement, les spectres infrarouges d'un grand nombre de composés organiques, à structures moléculaires plus ou moins complexes, sont, dans certaines firmes, répertoriés sur cartes mécanographiques). Cette carte d'identité est l'indispensable élément de comparaison en analyse quantitative. Ainsi, peut-être plus encore en infrarouge que dans d'autres domaines de λ, l'analyse spectrale fournira-t-elle de nombreux renseignements sur la constitution chimique des produits courants, qu'il s'agisse de l'identification des composants alimentaires, du contrôle de qualité des tabacs, de la concentration d'alcaloïdes dans un médicament ou de la tenue d'un pigment coloré dans une laque.

Si elles ont depuis longtemps été mises en œuvre dans de délicates opérations telles que le vieillissement artificiel des vins, la stabilisation des jus de fruits ou le séchage des peintures pour automobiles (1) et des denrées alimentaires, les modifications apportées dans la matière par le rayonnement infrarouge sont exploitées aussi dans des applications de plus en plus inattendues. En biologie, par exemple, citons la synthèse de la pénicilline, le diagnostic des inflammations à l'intérieur du corps humain, l'élevage des poussins sous la lampe infrarouge (accélération du développement des cellules).

Les photographies du *Christ en croix* du musée de Besançon montrent quelques aspects (v. p. 59) du chef-d'œuvre pris sous des éclairages variés. L'utilisation de l'infrarouge en photographie fournit un moyen d'identification puissant : les différents pigments des colorants ne réagissent pas de la même façon, toutes proportions gardées, en diffusion infrarouge qu'en lumière visible. Telle retouche habile, tel faux parfait seront décelés au moyen de longueurs d'onde supérieures au micron. De même, on a pu déchiffrer, au British Museum de Londres, des écrits égyptiens dont le temps avait réduit à ce point le contraste qu'ils étaient devenus illisibles. La littérature abonde en exploits accomplis grâce aux infrarouges.

téléphone et rayon rouge

Dès le début de ce chapitre, on a insisté sur ce caractère général des ondes électromagnétiques, « véhicules » de l'information. Abandonnons donc le domaine infrarouge avec un véhicule typique des communications : le téléphone. Il a été utilisé pendant la guerre, et son fonctionnement consistait en une modulation d'un faisceau infrarouge par le message à transmettre. Ce signal couvrait une certaine gamme d'audiofréquences. Il était traduit en variations d'amplitude reportées sur le faisceau émis de façon permanente par la source IR. Le récepteur effectuait l'opération réciproque. Il était ainsi possible de communiquer à l'insu de l'ennemi au moyen d'ondes qui échappaient aux détecteurs d'émissions radioélectriques.

Avons-nous ici un autre seuil à franchir? Nous nous sommes laissés glisser des micro-ondes hertziennes à l'infrarouge, qui nous est apparu comme un lien entre les ondes radio et le visible, entre les techniques des circuits accordés et de ce qui est à proprement parler la lumière, cette lumière dont on peut bien se demander si les ondes seront aussi facilement domesticables que leurs grandes sœurs lorsqu'il y aura un message à transporter. Le lecteur des années 60 trouve sans doute cette question bien superflue. Ne parle-t-on pas couramment des merveilleuses possibilités du laser? De ce pinceau de lumière rouge qui permet toute communication avec la lune ou un quelconque satellite? Il est de fait que les techniques de l'optique et celles de l'hertzien sont à ce point fondues dans le laser que l'on ne sait même plus quel langage adopter : celui des opticiens, ou celui des électroniciens (v. tome I).

où l'œil commence à y voir

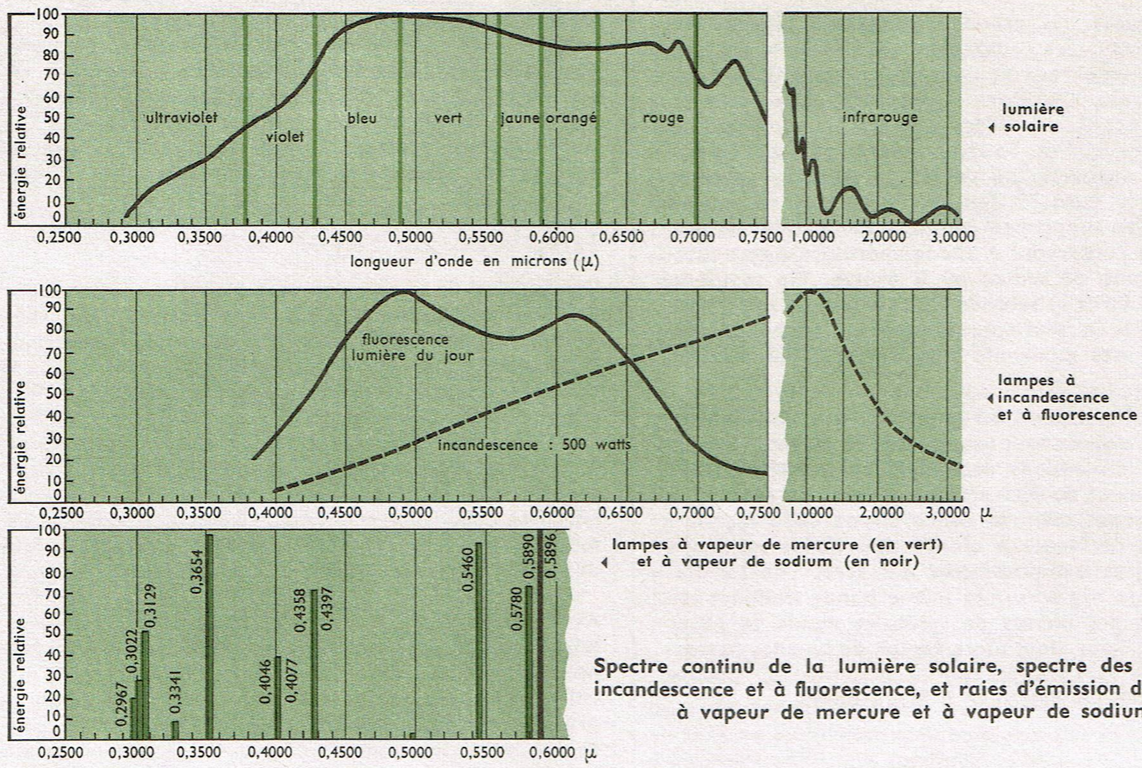
Le visible, c'est la gamme qui va du rouge au violet, de 0,8 à 0,4 µ. Les limites sont un peu floues et dépendent de l'œil. On pourrait presque dire que « le visible, c'est ce que découpent nos sens » (ce ne serait d'ailleurs pas tout à fait exact, puisque la peau ressent, par une impression de chaleur, une partie



Séchage de peinture à l'infrarouge.

(1) Il faut noter que ce genre d'opération n'est pas réalisé pour n'importe quelle longueur d'onde; ainsi la bande des 12 µ est fréquemment utilisée parce qu'elle correspond à une région spectrale où l'absorption sélective de l'eau (plus exactement du groupement OH) est importante.

Doc. Mazda.



Spectre continu de la lumière solaire, spectre des lampes à incandescence et à fluorescence, et raies d'émission des lampes à vapeur de mercure et à vapeur de sodium.

de l'infrarouge et même un faisceau hertzien très dense; la rétine est même sensible — si les ondes sont assez intenses — à la longueur d'onde de 1 μ, et l'on croit alors voir du rouge; la peau réagit aussi à l'ultraviolet par une sensation de brûlure). Les détecteurs du type de la cellule photo-électrique, les plaques et films photographiques sont largement utilisés dans le visible comme dans l'infrarouge ou dans l'ultraviolet.

Mais l'œil reste omniprésent. Le physicien dont le champ d'étude s'est longtemps cantonné au visible éprouve toujours quelque gêne, une certaine appréhension, lorsqu'il se transporte dans un autre domaine où il ne peut que faire confiance à l'appareil de mesure, sans « y mettre l'œil » directement.

L'un des chapitres suivants étant spécialement consacré à l'optique du visible, poursuivons notre descente, sortons de ce passage ruisselant de lumière, dans lequel il n'a pas toujours été possible de distinguer la limite entre deux tonalités : dans le spectre visible du Soleil, c'est-à-dire parmi diverses couleurs étalées au moyen d'un système approprié (prisme, disperseur quelconque), ou encore parmi les longueurs d'onde émises dans la région de sensibilité de l'œil humain, le lecteur pourrait-il déterminer avec précision le passage du jaune-orangé à l'orangé? Toutes les longueurs d'onde y sont présentes ou presque. On a affaire à un spectre continu. Au contraire, une lampe à décharge (comme celles qui éclairent quelques-unes de nos routes, les lampes au sodium par exemple) n'émet de lumière qu'en des régions spectrales de largeur réduite, les raies. Les deux raies jaunes ($\lambda_1 = 0{,}589$ μ et $\lambda_2 = 0{,}589\,6$ μ) étant les plus intenses dans le spectre du sodium, sa lumière paraît, en gros, jaune. La lumière de la lampe à mercure est à dominante bleue, les raies les plus intenses étant situées vers des longueurs d'onde plus faibles : vert ($\lambda = 0{,}546$ μ), bleu et violet (0,439 7 μ, 0,435 8 μ, 0,407 7 μ, 0,404 6 μ); celle du néon est rougeâtre, etc.

Après le violet, autour de $\lambda = 0{,}4$ μ, les ténèbres règnent de nouveau; nous n'y voyons presque plus.

l'ultraviolet

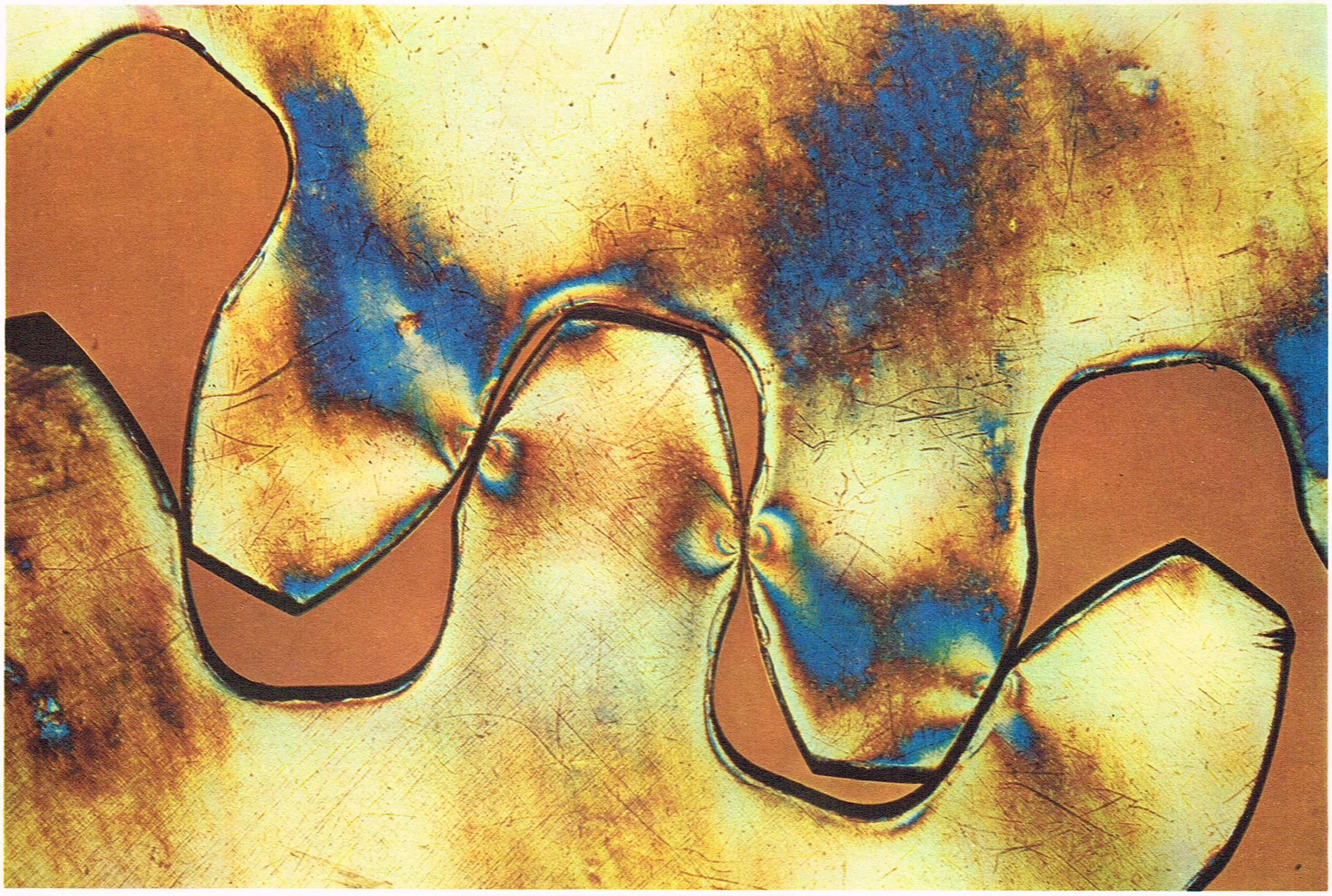
Comme pour l'infrarouge, on a souvent classé le rayonnement ultraviolet en proche (jusque vers 0,31 μ), moyen (0,28-0,25 μ) et lointain. Les sources? Il y a le Soleil bien sûr, et un grand nombre d'étoiles, mais l'atmosphère devient très vite à peu près opaque, à partir de 0,3 μ (ou 3 000 Å), par suite notamment de l'absorption par l'ozone (l'eau, elle, reste transparente au moins sous faible épaisseur jusqu'aux ondes de 0,18 μ). On en est donc réduit à produire l'ultraviolet « sur place », au moyen de lampes à électrodes métalliques entre lesquelles jaillit un arc électrique. Ces électrodes sont généralement entourées de matériaux transparents (silice, verres spéciaux) mais parfois on les utilise directement dans le vide.

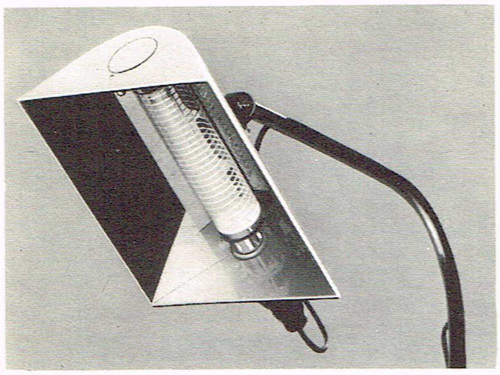


Ce timbre, très rare, n'est pas un faux, mais la photographie faite en ultraviolet moyen (0,25 μ) révèle qu'il a été habilement restauré.



Les amusantes correspondances entre l'art moderne — ci-contre, un détail d'une œuvre de Kandinsky — et une vision, bien concrète, d'un engrenage en photo-élasticimétrie.





Générateur de radiations ultraviolettes utilisé en thérapeutique.

L'action des radiations ultraviolettes sur la matière est très importante du fait des fréquences très élevées atteintes (que l'on compterait en milliards de mégahertz). En employant le langage qui deviendra familier d'ici quelques chapitres, le grain de lumière — le photon —, dont l'énergie individuelle est proportionnelle à la fréquence, arrachera sans difficulté des électrons à la matière irradiée (effet photo-électrique). Cela représente effectivement l'une des méthodes de détection les plus courantes. Le phénomène est, du même coup, à la base des multiples applications de l'ultraviolet.

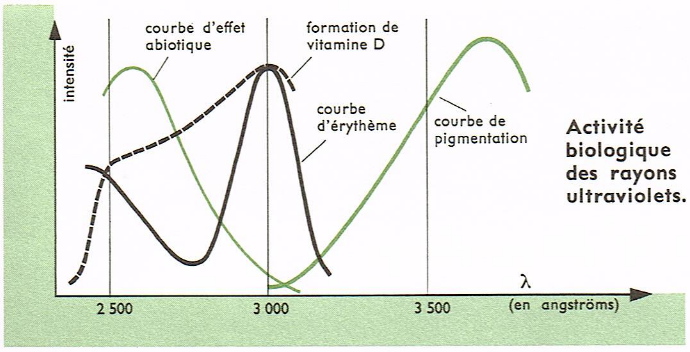
la fluorescence et ses surprises

Les lampes-tubes, dont les lumières de toutes couleurs inondent les rues et les façades, donnent une mine blafarde à la fraîche dactylo et d'étranges reflets au jambon accroché à l'étal du charcutier, mettent à profit la grande énergie ultraviolette dans le phénomène de la fluorescence. Ces lampes, souvent improprement nommées « tubes au néon » par les décorateurs, contiennent généralement du mercure, dont la vapeur, excitée par une décharge électrique à haute tension, émet une grande quantité d'ultraviolet. L'action de cet ultraviolet sur la substance dont on a tapissé l'intérieur de la lampe (sulfure de zinc fluorescent, par exemple) se traduit par une émission caractéristique, cette fois, de la substance choisie (blanc de luxe, lumière du jour, etc.).



Enseignes lumineuses à Broadway.

Doc. Philips Métalin et phot. U. S. I. S.



Activité biologique des rayons ultraviolets.

bain de soleil et huile de foie de morue

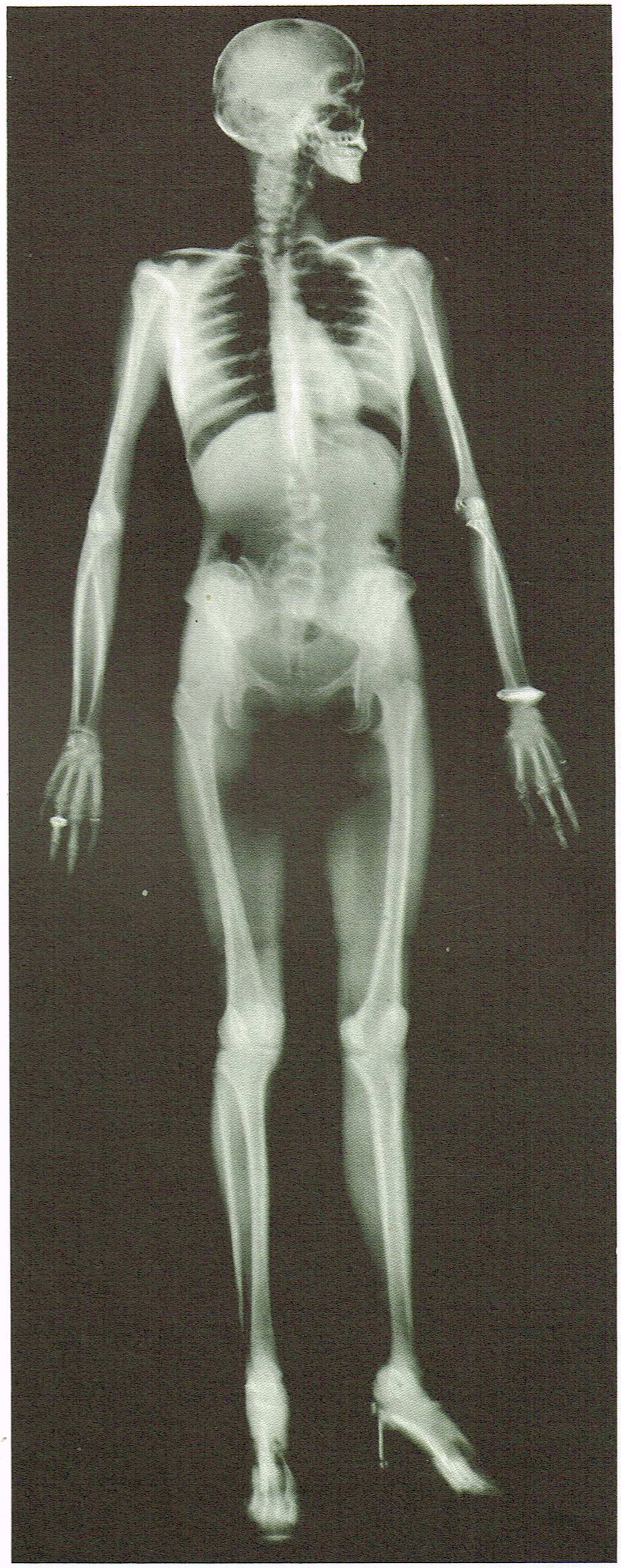
Chaque année, bien avant la période des vacances, commence la course au brunissage. Il faut sacrifier à la mode, faire comme tout le monde. Les crèmes et décoctions variées se répandent en même temps que les nombreux conseils et avertissements : « Brunissez vite et bien », « Attention aux terribles brûlures que peut causer le soleil! » Les courbes reproduites ici sont assez explicites à ce sujet. La région de pigmentation se situe bien avant la zone dangereuse. La limite se trouve au voisinage de $\lambda = 3\,200$ Å. Celle-ci correspondant à peu près au début d'un comportement particulier des couches profondes de la peau, l'érythème, réaction dans laquelle l'afflux de sang dans les capillaires tend à protéger le derme, l'exposition prolongée au rayonnement produit alors une irritation plus ou moins violente et peut conduire à une modification grave des tissus (brûlures au troisième degré). Alors demandez donc au boutiquier du coin une crème qui soit un filtre suffisamment sélectif pour ne rien laisser perdre des bons rayons, et qui supprime toute fréquence nocive. Mais il faudrait encore, pour que le soleil du mois d'août soit réellement bénéfique, qu'il reste assez d'ultraviolet pour qu'apparaisse dans l'organisme une dose de vitamines permettant de passer l'hiver. Peut-être serait-il plus sûr d'ingurgiter quelques cuillerées d'huile de foie de morue, de ce gros poisson qui en mange des plus petits, qui, eux, mangent le phytoplancton, ces algues marines microscopiques directement irradiées par l'ultraviolet solaire, usines à vitamines D.

Ainsi l'ultraviolet contribue à l'assimilation du calcium, à la formation d'agents vitaminiques antirachitiques, au cours de réactions chimiques dites « de photosynthèse » (réactions que l'on retrouve par exemple dans le durcissement des résines : la polymérisation), et, s'il peut devenir dangereux pour l'homme, il l'est fort heureusement aussi pour les bactéries, dont bien peu résistent aux longueurs d'onde voisines d'un quart de micron; la radiation $\lambda = 2\,537$ Å produite par un arc à mercure n'est-elle pas appelée la « radiation germicide »?

rayons X, photons et rayons gamma

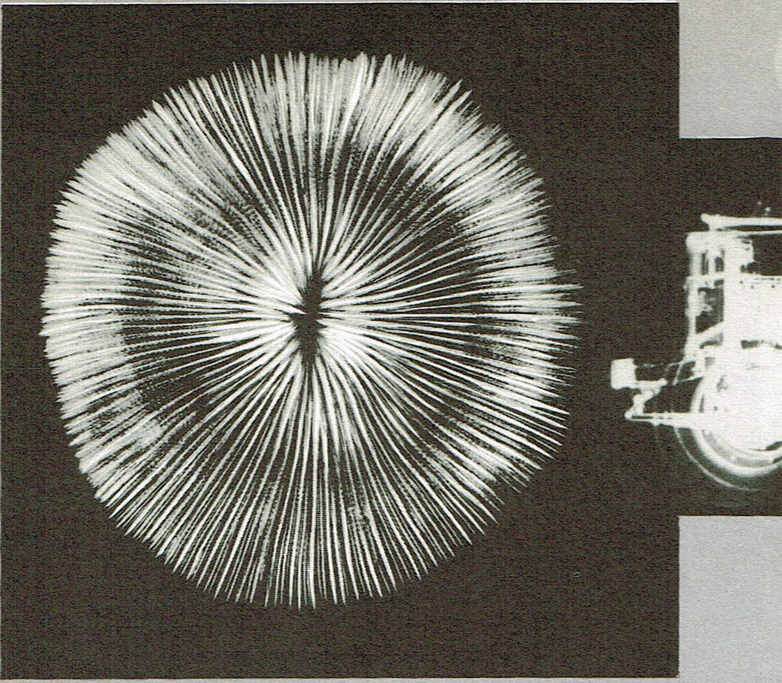
Lorsque sur chacune des divisions de notre échelle nous nous sommes arrêtés pour un instant, il a été difficile, chaque fois, de la quitter hâtivement tant elle apportait à notre existence de tous les jours. Laissons donc cet ultraviolet de l'assimilation chlorophyllienne, cet agent qui accélère les prises des greffes osseuses, ce moyen d'attendrissement du bifteck et de stérilisation des légumes. Abandonnons la lumière noire du cabaret aux ampoules violettes, qui, brutalement, brunit les plastrons immaculés, verdit les lèvres de nos compagnes, parsème de pourpre la meringue glacée, et poursuivons notre descente.

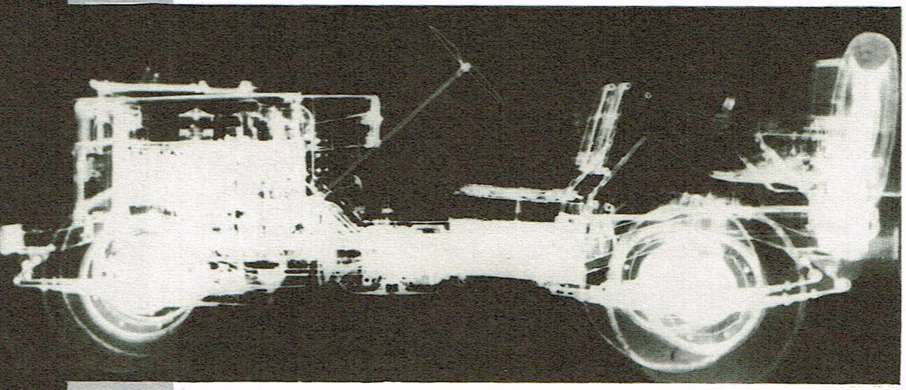
Voici les mystérieux rayons X, correspondant aux courtes longueurs d'onde (de quelques centaines, ou de quelques angströms). Il s'agit encore de rayonnement électromagnétique, mais leur mode de production, leurs propriétés s'expliqueraient beaucoup plus aisément si nous disposions de concepts qui



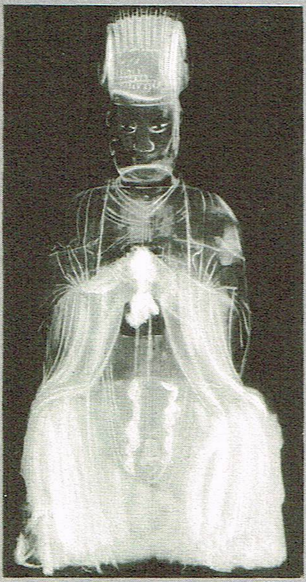
Radiographie d'une jeune femme.

Phot. Kodak.

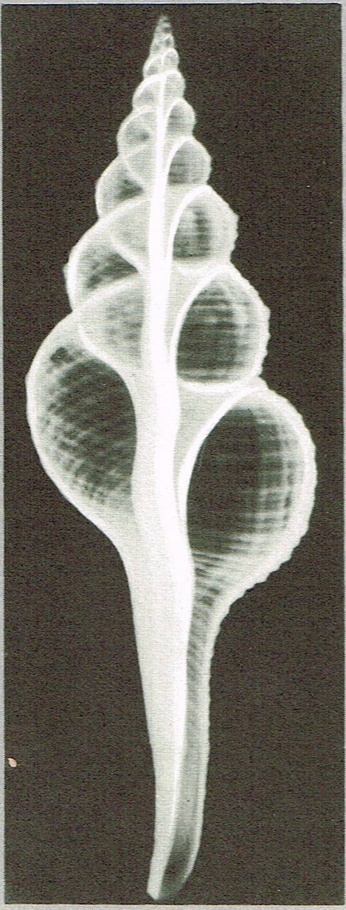


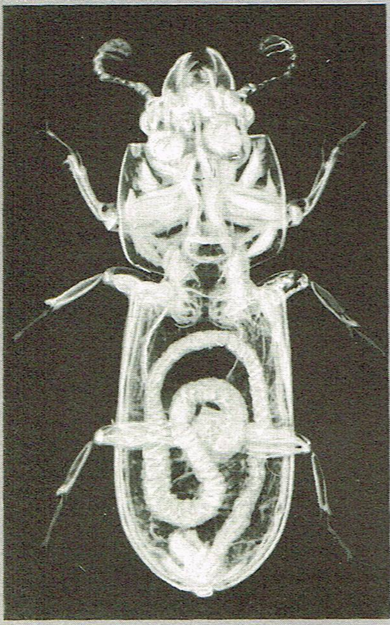


Radiographies d'objets divers : oursin, Jeep, statuette, fleur, insecte, coquillage.









seront exposés plus tard et auxquels il a déjà été fait allusion précédemment. Essayons cependant de schématiser la situation. On a vu que les collisions entre les molécules d'une substance donnaient lieu à une émission de rayonnement électromagnétique aux fréquences de l'infrarouge.

une prise de contact avec les photons

Le fait de parler de grains de lumière, ou, plus généralement, de concentrations d'énergie électromagnétique, ne représente pas quelque chose d'inconcevable en soi. Nous avons dit aussi que l'énergie de ces grains, les photons, était proportionnelle à la fréquence de l'onde qui leur est ainsi associée.

Dans le visible, les énergies sont plus grandes, puisque la fréquence est plus élevée. Ce ne sont plus seulement les molécules qui entrent en jeu, mais les électrons tournant autour du noyau atomique. Puisqu'il s'agit de particules chargées, le mouvement crée un champ $\vec{E}$ et, finalement, une onde électromagnétique dont on peut calculer aisément la fréquence : à un électron tournant autour d'un noyau sur une orbite de 1 Å de diamètre correspondrait l'émission d'une onde de 10^{15} Hz environ, c'est-à-dire de fréquence optique.

les rayons X

Les rayons X, eux, prennent naissance lors de la décélération d'un flot d'électrons circulant à grande vitesse et rencontrant une plaque cible de tungstène, par exemple. L'onde de freinage, la radiation X, présente encore les propriétés déjà rencontrées, mais si la fréquence n'est pas trop élevée, la longueur est tout à fait comparable aux dimensions habituelles des réseaux cristallins. Dès lors, les rayons X représenteront l'un des outils les plus souples de l'exploration de la matière (étude des cristaux, des couches superficielles métalliques, etc.).

Nous nous contenterons ici d'admirer quelques photographies, telles les reproductions de microradiographies obtenues au moyen de rayons X de plus ou moins grande énergie, c'est-à-dire, en gros, plus ou moins pénétrants. D'autre part, chacun de nous a eu au moins une fois l'occasion de « passer une radio »; cette transparence aux rayons X, nous la retrouvons sur les clichés.

les rayons gamma

Enfin, vers l'extrémité du spectre, le rayonnement gamma, plus pénétrant encore, véhiculant de très grandes énergies, de fréquence de l'ordre de 10^{21} Hz (c'est-à-dire mille milliards de milliards de cycles par seconde, ou encore de 0,01 à 1,5 Å de longueur d'onde), reste, par ses propriétés, en diffraction par exemple, un élément de la famille électromagnétique.

Phot. Kodak.

Et plus on s'en va vers des fréquences élevées, plus le rayonnement se présente sous forme corpusculaire et perd son aspect ondulatoire. Nous arrivons aux rayons cosmiques, qui sont en très grande majorité des noyaux d'atome au mouvement très rapide. Parmi eux se trouvent quelques rayons X ou gamma de grande énergie qui appartiennent au spectre électromagnétique; on les détecte toujours par leur effet corpusculaire : effet Compton, création de paires d'électrons, etc.

●

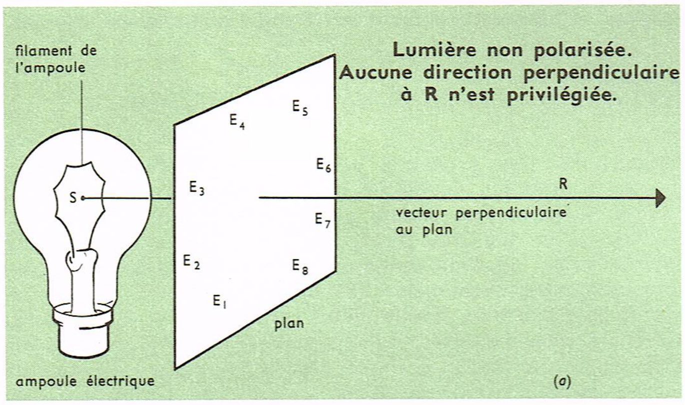
Voici fini notre survol des longueurs d'onde. Il nous a permis de reconnaître une étonnante unité à travers des domaines aussi variés que télescopes et radiotélescopes, émissions radio et télévision, tubes à cavité et klystron, radar et cuisinières à magnétron, fer à repasser et photo-détecteurs, évaporographes et systèmes de Golay, tubes à fluorescence et rayons X, photons et rayons gamma.

Nous avons laissé de côté l'une des gammes les plus importantes des longueurs d'onde (les ondes visibles, de 0,4 à 0,8 μ), et la science correspondante : l'optique. Ou, plutôt, les sciences correspondantes, car il y a plusieurs chapitres de l'« optique ».

Avant de lui donner une place d'honneur, qui lui revient par la primauté de l'œil sur tout autre instrument d'observation et de mesure, nous allons faire quelques excursions dans des domaines aussi riches de signification théorique que d'applications : la polarisation, l'analyse harmonique, les résonances et l'effet Doppler-Fizeau nous ouvriront les portes de la physique moderne.

LA LUMIÈRE POLARISÉE

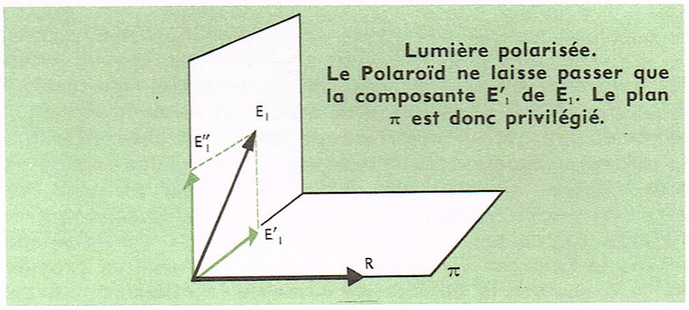
Un filament S porté à l'incandescence émet un rayonnement électromagnétique, qui nous parvient le long du rayon SR. En fait, chaque point du filament est une petite source, un émetteur élémentaire, un doublet, si l'on veut bien reprendre la terminologie de l'électromagnétisme. L'orientation de chacun de ces doublets, dont le nombre est immense, n'a aucune raison de s'aligner sur celle du voisin. Les vecteurs $\vec{E}$ possèdent donc toutes les orientations possibles dans le plan perpendiculaire à SR, et la répartition des orientations s'effectue selon les lois du hasard. En général, donc, la source rayonne de la lumière sans qu'aucune orientation soit privilégiée pour les vecteurs électriques élémentaires, nullement soumis à une sélection au départ de l'émetteur. On dit que la source « rayonne de la lumière naturelle ».



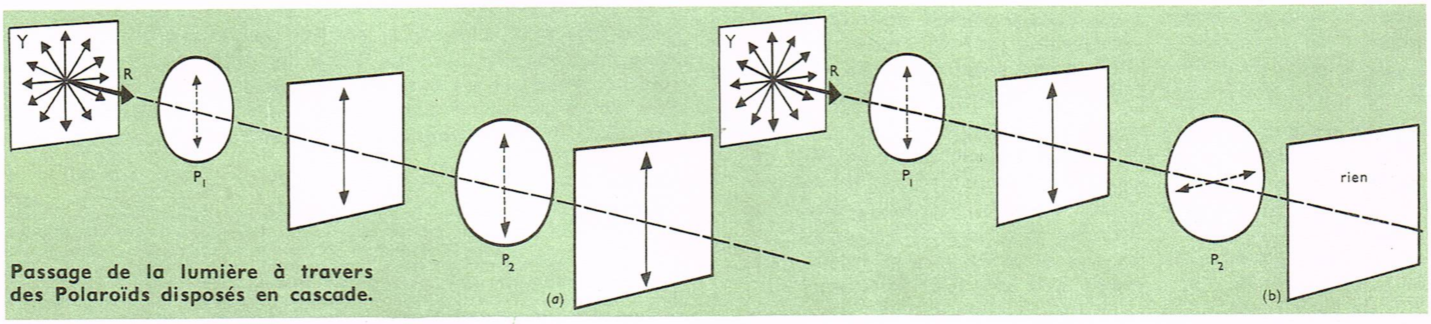
Lumière non polarisée. Aucune direction perpendiculaire à R n'est privilégiée.

un filtre à vecteur : le Polaroïd

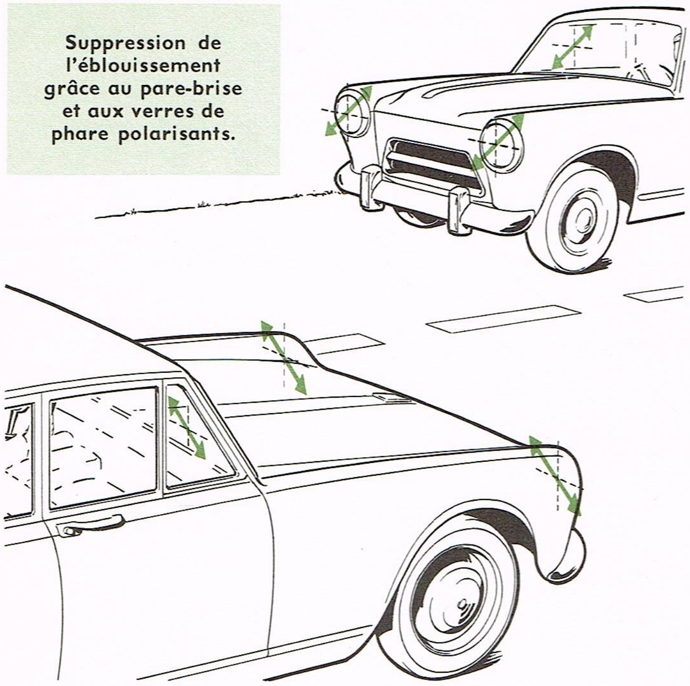
Regardons maintenant la lampe à travers un Polaroïd P_1. Apparemment, rien ne se passe : on voit la lampe; il y a peut-être un peu moins de lumière, c'est tout. Plaçons un second Polaroïd P_2 à la suite du premier, et faisons-le tourner dans son plan. À travers l'ensemble l'intensité varie. Pour un certain angle de rotation, l'extinction est même quasi totale. Comment s'expliquer le phénomène? D'abord, si besoin en était encore, cette extinction montrerait que le vecteur de Fresnel ne peut être longitudinal puisque alors il ne pourrait être réduit par rotation. Quel a donc été le rôle de P_1? Le premier Polaroïd a effectué un triage. Il n'a laissé passer, du paquet de flèches du plan Y qui représentent les diverses orientations de $\vec{E}$, que leurs composantes selon une seule direction, celle que nous avons représentée en pointillés sur P_1. On dit qu'il a polarisé la lumière, que c'est un polariseur. L'onde est polarisée rectilignement, $\vec{E}$ garde une orientation fixe et oscille uniquement le long de cet axe particulier. La lumière naturelle n'est pas polarisée. Le second Polaroïd P_2, identique au premier, va lui aussi effectuer un triage pour son propre compte et ne laisser passer que les vibrations pour lesquelles le vecteur $\vec{E}$ sera orienté parallèlement à la direction repérée en pointillés sur P_2. Ainsi aura-t-on en *a* un maximum de lumière transmise par P_2. En *b* aucune transmission n'est plus effectuée, les Polaroïds sont dits « croisés » (l'énergie lumineuse disparue a finalement servi à ... chauffer le Polaroïd). Ils étaient « parallèles » en *a*.



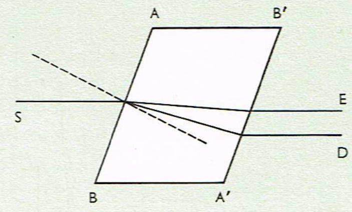
Lumière polarisée. Le Polaroïd ne laisse passer que la composante E'_1 de E_1. Le plan π est donc privilégié.



Passage de la lumière à travers des Polaroïds disposés en cascade.



Suppression de l'éblouissement grâce au pare-brise et aux verres de phare polarisants.



Biréfringence.
Le rayon incident S fournit deux rayons réfractés qui émergent parallèlement en D et E de la lame biréfringente ABA'B'.

Ce qui précède est si peu abstrait, en définitive, que la meilleure façon de se rendre compte des choses est de faire l'expérience en utilisant les Polaroïds, que l'on trouve désormais dans le commerce. Quitte à devoir reprendre plus tard une explication plus détaillée, on peut fort bien expérimenter d'abord. Par exemple, regardez à travers un Polaroïd la lumière qui se réfléchit sur un meuble brillant; son intensité varie lorsqu'on tourne le Polaroïd. C'est que le meuble a effectué un premier tri, il a polarisé la lumière au cours de la réflexion. Suivant son orientation, le Polaroïd laisse passer plus ou moins de lumière. Le comportement du bois verni, lors de la réflexion, est fonction finalement de l'orientation des vecteurs de Fresnel élémentaires du faisceau lumineux incident. Les propriétés de polarisation du meuble permettent d'éviter toute la lumière qui voilait les détails et empêchait les veines du bois de « ressortir ».

Quel effet le photographe ne tirera-t-il pas de ce phénomène, rien qu'en ajustant devant l'objectif de son appareil une feuille de Polaroïd qu'il oriente à loisir!

Regardons le ciel juste au-dessus de nous, puis très bas vers l'horizon à travers notre même Polaroïd; ce bleu si pur que l'on eût dit si naturel n'est-il pas polarisé différemment à l'horizon et au zénith?

les pare-brise antiphares

Puisqu'une orientation convenable d'un polariseur permet de réduire la lumière déjà polarisée au moins partiellement, comment ne pas mettre à profit cette propriété pour tenter de limiter l'éblouissement du conducteur sur la route, le soir, lorsque le soleil rasant fait briller la chaussée? C'est à ce moment plus qu'à tout autre que les lunettes polaroïdes sont bien utiles.

Et, pendant que nous y sommes, ne pourrait-on imaginer un système « antiphares » qui éviterait aux automobilistes de s'éblouir mutuellement la nuit? Un ou deux schémas paraîtront plus explicites que tout long commentaire. Supposons que, très judicieusement, les conducteurs des voitures A et B aient recouvert de feuilles polaroïdes les phares et les pare-brise de leurs véhicules respectifs. On a même pris soin d'orienter les Polaroïds de telle façon qu'ils ne laissent passer que les vibrations dont la direction est inclinée à 45° de la gauche vers la droite (l'opération étant dirigée par le conducteur depuis son volant). Les deux voitures arrivent l'une vers l'autre. Le pare-brise de chacune ne laissant pratiquement passer les vibrations que suivant la direction qui se trouve être perpendiculaire à celle qui est employée par le phare de l'autre, tout éblouissement sera évité. Mais la lumière de ses propres phares, chaque conducteur la verra puisque son pare-brise et ses phares sont « parallèles ». Il repérera donc les obstacles éclairés par ses phares; sûrement aussi un peu de lumière de son vis-à-vis, car les Polaroïds et leur montage ne seront pas tout à fait réglés, et il faudra tenir compte aussi de la largeur et de l'inclinaison des faisceaux. Un tel système n'est cependant pas dépourvu de quelques inconvénients sérieux, ne fût-ce que son prix élevé, et malheureusement, le fait est que sa réalisation sur toutes les voitures semble actuellement encore du domaine de l'utopie.

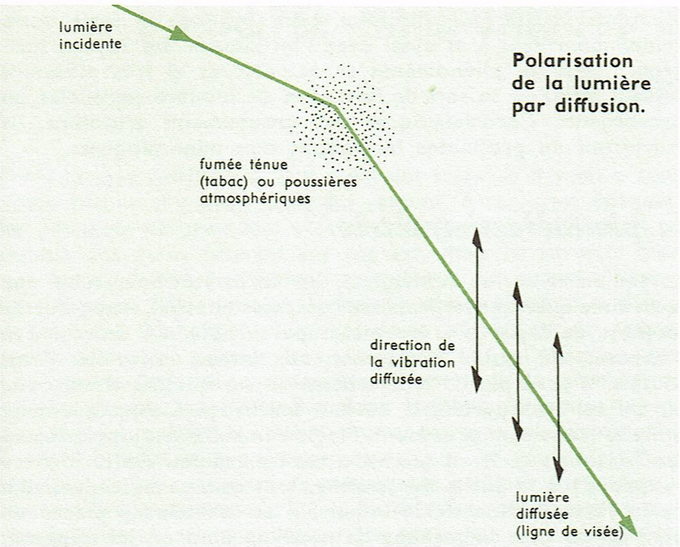
biréfringence et double réfraction

Il y a plus de trois cents ans, un certain M. Bartholinus s'aperçut qu'un morceau de calcite (spath d'Islande) dédoublait les images. Lorsque l'on envoie un faisceau lumineux étroit sur un bloc de ce matériau transparent, on observe le dédoublement suivant : un premier rayon, dit « rayon ordinaire », suit normalement le chemin que l'on aurait pu prévoir d'après les lois de la réfraction, tandis qu'un second rayon, « extraordinaire », semble défier les règles familières. De plus, on constate à la sortie du cristal que ces rayons sont polarisés suivant deux directions perpendiculaires, ce que l'on vérifie aisément en faisant tourner un Polaroïd dans un plan perpendiculaire à ces rayons émergents.

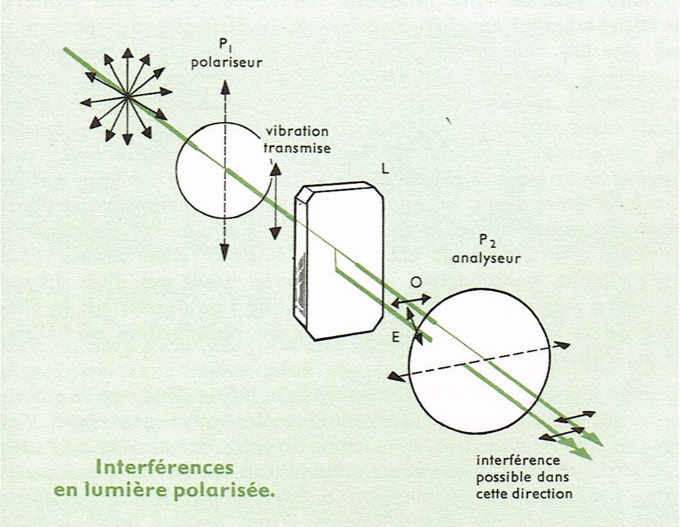
Comment expliquer l'apparition de cette double réfraction? Quand la lumière naturelle, non polarisée, se propage dans un milieu isotrope (c'est-à-dire dont les propriétés sont les mêmes dans toutes les directions), les lois « habituelles » s'appliquent normalement. Mais le spath est anisotrope : seule la lumière dont les vibrations sont parallèles à l'une ou l'autre de deux directions perpendiculaires est transmise, et les deux catégories de vibrations se propagent à des vitesses différentes.

Voici donc une manière de produire de la lumière polarisée. Il y avait déjà cet autre moyen de triage que représente la réflexion vitreuse (le faisceau réfléchi était polarisé, au moins partiellement). Nous connaissons les Polaroïds. (1) Peut-être même les avons-nous « expérimentés » au cinéma en relief. Il est bien d'autres sources encore de lumière polarisée. Ainsi, le ciel très bleu paraît plus sombre par endroits, suivant certaines orientations, lorsque l'œil est muni de lunettes polaroïdes : c'est que les fines gouttelettes de brume, responsables déjà de la couleur bleue (l'intensité de la lumière diffuse est inversement proportionnelle à la quatrième puissance de la

(1) Le Polaroïd est constitué de molécules organiques étirées, orientées suivant une direction particulière. Les molécules, par exemple d'hérapatite, d'iodure de kératine, ou d'iodosulfate de quinine en solution dans un milieu volatil éliminé après orientation des microcristaux. Actuellement, on étire des feuilles de polyvinyle; l'orientation des molécules produit une double réfraction, puis l'élimination de l'une des deux vibrations perpendiculaires est obtenue (par absorption sélective, ou dichroïsme) en trempant les feuilles dans une solution d'iodine.



Polarisation de la lumière par diffusion.



Interférences en lumière polarisée.

(L) est une lame cristalline d'axe optique perpendiculaire à la direction de propagation et incliné sur la direction de la vibration incidente (l'axe optique est la direction suivant laquelle aucun phénomène particulier ne se produirait : le comportement du milieu est alors celui d'un corps isotrope; cet axe optique correspond à la symétrie d'ordre le plus élevé). Ici, l'axe optique étant placé dans un plan perpendiculaire à la direction de propagation, il y a biréfringence.

longueur d'onde; donc le bleu, dont la longueur d'onde est inférieure à celle du rouge, subit une plus grande diffusion), polarisent aussi la lumière. La vibration prépondérante est perpendiculaire au plan que forment le rayon de lumière incidente et la ligne de visée.

Les phénomènes précédents apparaissent très différents les uns des autres. En fait, ils traduisent une constante interaction de la lumière et de la matière, et par là même, par le comportement de cette matière sous l'effet d'une vibration particulière, ils mettent en évidence certaines propriétés intimes de la matière.

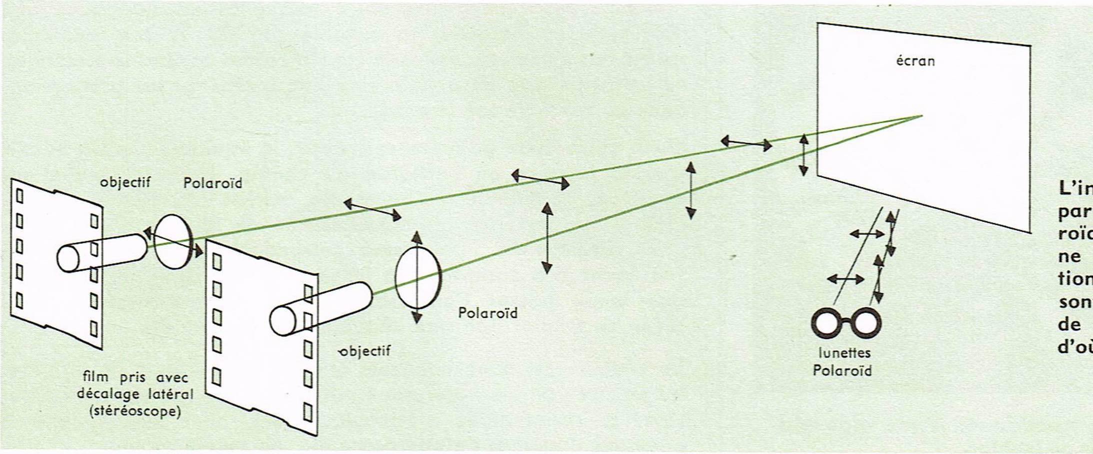
Pasteur et la polarisation rotatoire

Ne quittons pas ce domaine, où nous voyons combien sensible est le vecteur lumière à la structure du milieu qui le propage, sans mentionner la polarisation rotatoire : l'orientation du plan de vibration de la lumière qui traverse une dilution d'acide lactique (ou encore un barreau de quartz, que l'on a taillé de telle manière qu'il ne présente aucune biréfringence pour le faisceau incident) varie proportionnellement à l'épaisseur traversée.

Pasteur montrait en 1856 qu'un défaut de symétrie d'un cristal était susceptible de faire tourner le plan de polarisation de la lumière qui le traversait; il jetait ainsi les bases de la **stéréochimie.** Du fait de l'asymétrie de l'atome de carbone, certains corps organiques de formules identiques existent suivant plusieurs variétés possédant des propriétés physiques, biologiques ou chimiques différentes. C'est le cas, par exemple, du glucose ou de l'acide tartrique. Chacun de ces corps se présente sous différentes formes (isomères stéréochimiques); les solutions aqueuses de ces substances sont optiquement actives; l'angle dont elles font tourner le plan de vibration d'un faisceau lumineux qui les traverse permet de leur attribuer un pouvoir rotatoire spécifique et, par là même, de les identifier.

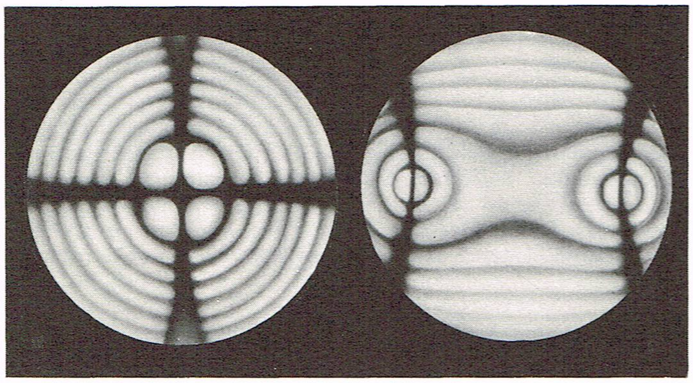
interférence et polarisation chromatique

Nous verrons par ailleurs qu'une lame cristalline biréfringente peut donner lieu à des phénomènes d'interférence entre les vibrations extraordinaire et ordinaire. Ces vibrations étant orthogonales, on place un second Polaroïd P_2 à la sortie de la lame, de manière à ne transmettre que des composantes parallèles entre elles (si l'on se contentait d'examiner la lumière à la sortie de la lame, sans interposer d'analyseur, la vibration résultant de la composition E et O serait elliptique). De plus, il existe une relation de phase bien définie entre ces vibrations (cheminant à l'intérieur du cristal à des vitesses différentes, elles présentent un certain retard l'une sur l'autre lors de l'émergence). Les deux vibrations peuvent donc interférer.

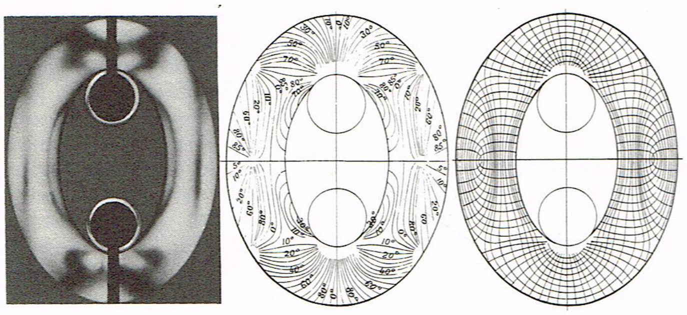


Lunettes en Polaroïd et cinéma en relief.

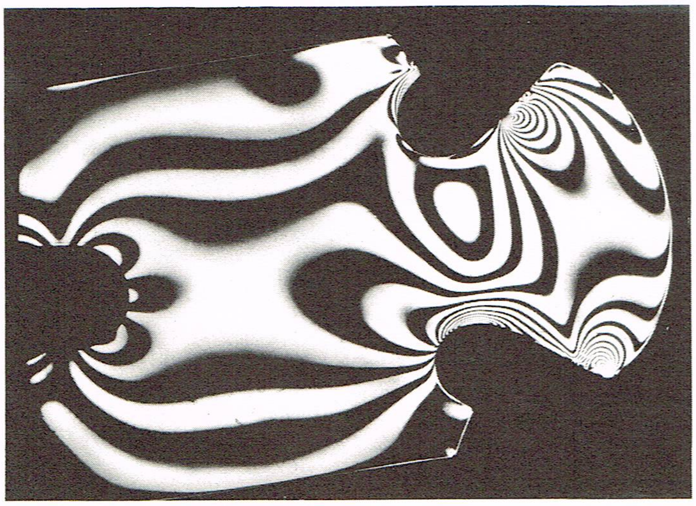
L'image de droite n'est vue que par l'œil droit, puisque le Polaroïd placé devant l'œil gauche ne laisse passer que les vibrations orthogonales à celles qui sont projetées par le système de droite, et réciproquement : d'où l'évidente restitution stéréoscopique.



Franges d'interférence fournies par des lames minces vues au microscope polarisant. *A gauche*, lame uniaxe; *à droite*, lame biaxe.



Étude photoélasticimétrique des efforts dans un maillon de chaîne. *A gauche*, photographie du modèle en Celluloïd vu dans un faisceau de lumière parallèle. *Au milieu*, réseau de lignes isoclines le long desquelles se forme une traînée sombre pour chaque orientation donnée des lames polarisantes. *A droite*, lignes isostatiques, se coupant à angle droit, dont les directions sont celles des contraintes.



Photoélasticimétrie : réseau d'isochromes d'une maquette de pied d'aube de turbine.

Le retard relatif des vibrations à la sortie de la lame varie évidemment avec λ et aussi avec l'inclinaison des rayons incidents. Ainsi les phénomènes seront colorés et très différents suivant que l'on se sert de faisceaux de lumière parallèles ou convergents. Caractéristiques des groupements cristallins, ils fourniront de précieuses indications aux minéralogistes.

la photoélasticimétrie

Lorsqu'un matériau homogène, un corps isotrope, subit une contrainte quelconque (traction, compression, etc.), il se déforme toujours de façon plus ou moins perceptible, en cherchant à s'opposer, au moins localement, aux forces auxquelles il est soumis. Il perd par la même occasion ses qualités d'isotropie, et, s'il est transparent, il devient biréfringent. Appliquons-lui alors le traitement précédent. Plaçons-le entre deux polariseurs croisés tandis qu'il est soumis à ces contraintes. De la lumière réapparaît à la sortie du système, tout comme précédemment après l'introduction de la lame. S'il est difficile de placer un pont ou un gros engrenage de machine dans un tel dispositif, on peut réaliser une maquette transparente en une matière particulièrement sensible aux efforts, le Plexiglas par exemple, qui devient vite biréfringent pour peu qu'on le presse. Une « carte » des tensions obtenue de cette façon sera aisément interprétée qualitativement et quantitativement. Cette carte, c'est aussi celle que l'on aperçoit très bien lorsqu'on examine les pare-brise des voitures avec des lunettes polaroïdes, ou parfois même sans lunette du tout, mais sous un certain angle. On peut dire, alors, que la lumière du ciel, partiellement polarisée, traverse la région du pare-brise à l'intérieur de laquelle ont été artificiellement créées des tensions déterminées (qui permettront, en principe, le découpage assez net d'un disque de verre lors d'un choc éventuel) et que l'analyseur est la surface de la vitre elle-même. Le taux de polarisation est faible, et les couleurs pâles.

Et puisqu'il a été fait allusion aux ponts, il faut indiquer l'emploi, dans les techniques de surveillance des ouvrages d'art, des jauges d'extensométrie à mesure directe, constituées par des réglettes en plastique recouvertes d'une feuille de polariseur. Elles sont ajustées de manière à faire apparaître un système d'interférence : les déplacements d'une frange noire autour d'une position repère renseigne sur les contraintes subies par la jauge et, par conséquent, sur les déformations de l'ouvrage au cours du temps.

effet Kerr et obturateurs ultra-rapides

Un autre exemple de biréfringence accidentelle est fourni par l'effet Kerr : à l'intérieur d'un bac de petites dimensions (cellule de l'ordre de quelques centimètres) rempli de nitrobenzène, plongent deux électrodes entre lesquelles est appliquée une différence de potentiel importante (20 000 V, par exemple) au moyen d'une impulsion de courte durée. Le champ électrique résultant produit dans le liquide une biréfringence proportionnelle au carré de son intensité.

Placé entre deux polariseurs croisés, le liquide ne laisse passer aucune lumière en l'absence du champ. Lorsque celui-ci est appliqué, le liquide se comporte comme un cristal uniaxe, c'est-à-dire qu'il est alors susceptible de fournir deux vibrations émergentes, polarisées perpendiculairement. Celles-ci possèdent des composantes suivant la direction de vibration que laisse passer l'analyseur. Donc, la biréfringence ainsi créée se traduit par une réapparition de lumière.

Le système est donc utilisable comme modulateur de lumière, ou encore comme obturateur ultra-rapide, dont la durée d'ouverture, commandée électroniquement, peut être réduite à quelques dixièmes de microseconde, ou même moins.

Phot. O. N. E. R. A.

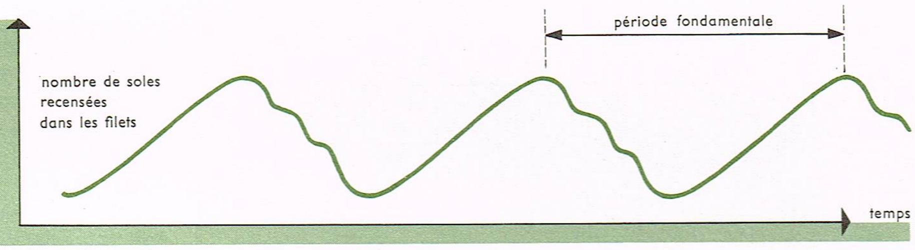
L'ANALYSE HARMONIQUE; UNE THÉORIE? UNE TECHNIQUE? DU DÉCOUPAGE...

Dans la mer Adriatique, il y avait des soles — il y en a sans doute encore. Il y avait aussi des requins. A certaines époques, les pêcheurs ne prenaient plus de soles dans leurs filets. Il ne semble pas qu'ils prissent des requins, mais ils savaient bien que la population des voraces augmentait. Le savant Volterra fut chargé au siècle dernier de fournir une explication, à défaut de remède. D'abord il s'aperçut de la périodicité du phénomène. Calculer, prévoir le cycle, indiquer à l'avance, pour une époque donnée, les courbes d'évolution du nombre de soles et de requins, c'était lier deux facteurs, déterminer entre eux de façon précise une relation de cause (appétit des requins) à effet (diminution des chances de longue vie des soles). Pour satisfaire le grand appétit de ces requins groupés dans l'Adriatique, il faut beaucoup de soles. Si les soles sont mangées plus vite qu'elles ne se reproduisent (ce qui est le cas, puisque les reproducteurs disparaissent), leur nombre diminue. La pitance s'amenuise. Des requins affamés, seuls les plus forts subsistent, leur nombre diminue. (S'ils choisissent de s'entre-dévorer, leur disparition sera même plus rapide encore.) Alors les soles, moins tourmentées, se repeuplent tranquillement, fournissant aux requins rescapés une nourriture disponible de plus en plus abondante. Le nombre des requins recommence à croître. Les soles seront mangées. Le processus se reproduit.

Finalement, le nombre de soles ne pourrait-il se maintenir stationnaire? En d'autres termes, l'équilibre peut-il être atteint? Certainement pas, puisque l'évolution des requins « suit » l'évolution des soles avec un certain retard, un « déphasage » entre les fluctuations correspondantes.

La corrélation est suffisante entre les accroissements et les diminutions des populations de l'un et l'autre groupe pour que le phénomène global apparaisse périodique. Mais non pas simple. Le graphique indiquant le nombre de soles, ou de requins, au cours du temps, n'a aucune raison de montrer un amortissement, et la courbe, pas plus, d'être sinusoïdale. La prévision et l'explication des faits deviendraient élémentaires si l'on connaissait exactement, à chaque époque, les populations respectives. Or, le seul paramètre effectivement déterminable, de manière sans doute approximative, est le nombre de soles recueillies dans les filets des pêcheurs. Supposons que la courbe correspondante, fonction du temps, suive le tracé, encore relativement peu complexe, de la figure ci-dessous. Les bosses et les creux peuvent fort bien être dus à l'état de la mer ou des vents, défavorable en certaines époques de l'année. Mais on ne doit guère s'attendre que l'intervalle de temps séparant les tempêtes de l'Adriatique — la période moyenne des tempêtes — soit du même ordre de grandeur que la période de repeuplement des requins ou des soles.

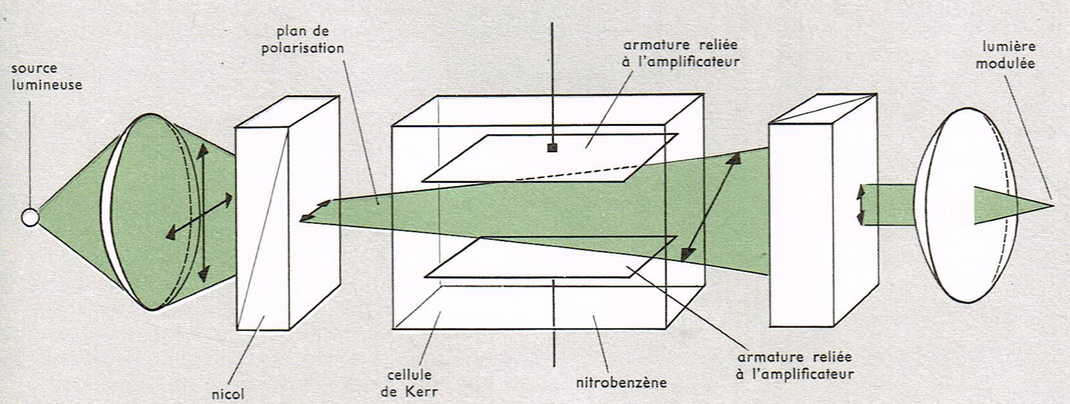
Enregistrement d'un phénomène périodique : la variation du nombre de soles pêchées dans la mer Adriatique.



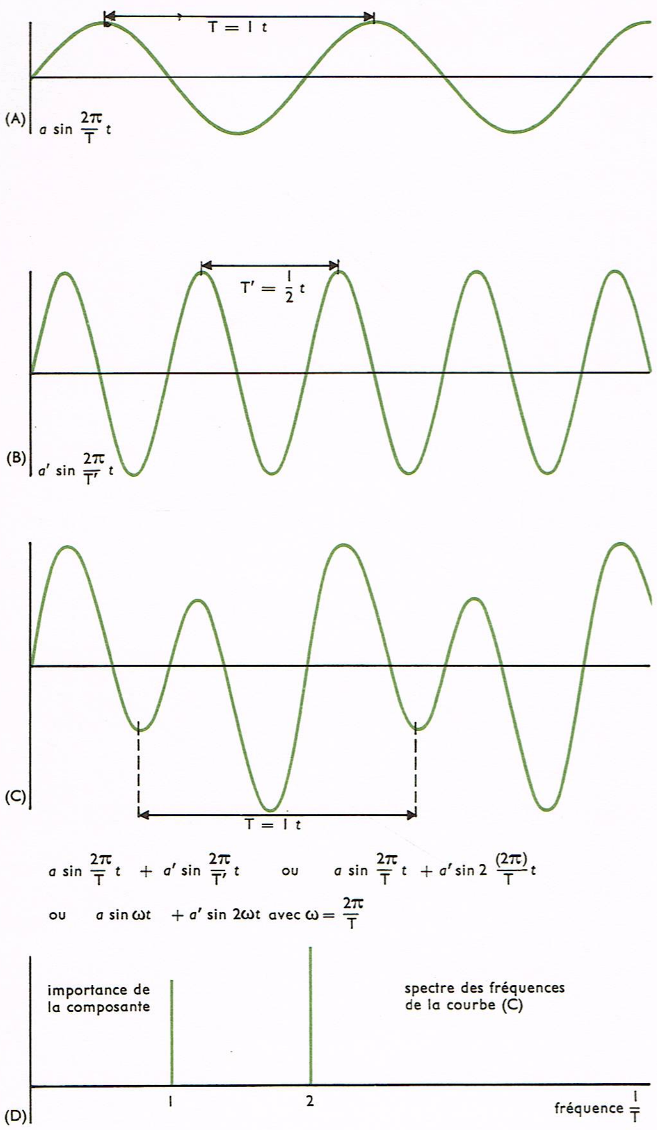
Ainsi, le problème se ramène à chercher en premier lieu les « périodicités cachées » que la courbe résultante ne met pas en évidence. Il se peut aussi que certains événements élémentaires présentent une fréquence de répétition (inverse de la période la plus nette observable) qui soit double ou triple de celle du phénomène général, c'est-à-dire de la fréquence fondamentale. L'opération à effectuer est une analyse harmonique.

une analyse unique, diverses applications

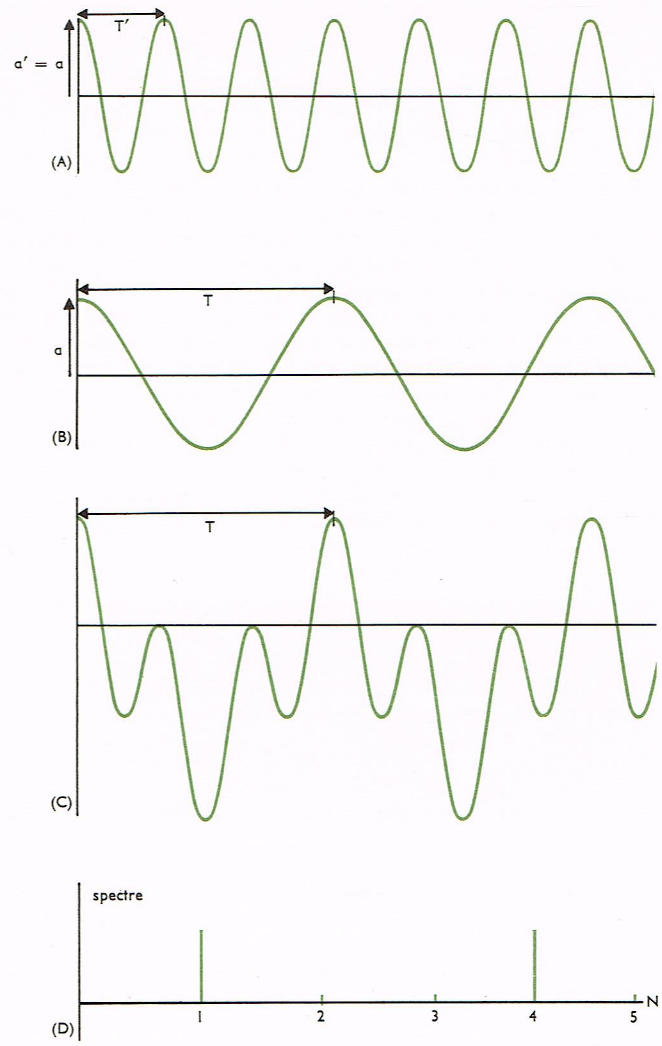
Illustrons par des exemples simplifiés la technique de cette opération si utile à beaucoup de sciences et si commune à beaucoup de disciplines (analyse du cycle solaire; météorologie; analyse factorielle, etc.). Un phénomène périodique élémentaire de période T est représenté par une sinusoïde A, et un second phénomène périodique élémentaire de période T′, par la sinusoïde B. La superposition des deux phénomènes, obtenue en additionnant les ordonnées de A et les ordonnées de B à



Modulation de la lumière avec la cellule de Kerr. Une variante de ce dispositif est utilisée pour la sonorisation des films cinématographiques.



La courbe (C) résulte de la superposition des deux phénomènes périodiques différents (A) et (B). Le spectre des fréquences de (C) permettra de distinguer les deux composantes.



Modulation d'un mouvement périodique (A) par un autre mouvement périodique (B) de fréquence inférieure. (C) est la courbe résultante et (D) son spectre.

chaque instant, conduira à la courbe C, dont la période, dans notre cas, est encore T.

Imaginons, maintenant, qu'au lieu d'avoir à effectuer une synthèse de C nous nous soyons trouvés devant la courbe C représentant un phénomène à interpréter, dont on aurait aperçu une périodicité, mais qui n'aurait pas été tout à fait simple. L'analyse harmonique aurait consisté à déterminer les composantes de C, c'est-à-dire à trouver leurs périodes et l'importance relative de chacune de ces composantes — importance donnée par les amplitudes a et a'. On a coutume de représenter ces composantes en portant en abscisses la période ou fréquence de chacune d'elles et, parallèlement à l'axe des ordonnées, des segments de longueurs proportionnelles à l'importance de la composante. Un tel graphique est appelé **spectre du phénomène** représenté par C. Dans le cas présent, nous avons deux fréquences : $N = \frac{1}{T}$ et $N' = \frac{1}{T'}$. La fréquence $N' = \frac{1}{T'}$ est double de la fréquence $N = \frac{1}{T}$ (puisque $T' = \frac{1}{2}T$). On dira que le mouvement B est le premier **harmonique** du mouvement A, celui-ci étant le **fondamental.**

Nous avons affaire à un phénomène particulièrement simple; le spectre ne comporte que deux termes : le fondamental et le premier harmonique.

Dans le second exemple d'analyse, la courbe C comporte encore deux composantes : le terme fondamental A de période T et l'harmonique d'ordre 4, de période $T' = \frac{T}{4}$, c'est-à-dire de fréquence $N' = 4N$. Ces deux termes étant de même importance, les deux termes du spectre sont représentés par des segments de même longueur.

Les spectres précédents sont discontinus; ils comportent un nombre fini ou infini de termes dont les fréquences sont toutes multiples de la fondamentale, ce qui revient à dire que le mouvement analysé est constitué d'une série unique de termes harmoniques. Nous allons rencontrer des mouvements dans lesquels un très grand nombre de fréquences fondamentales sont présentes simultanément, chacune de ces fréquences pouvant être accompagnée ou non de ses harmoniques. Lorsque le nombre des mouvements élémentaires nécessaires à la synthèse du mouvement original (et notamment chaque fois qu'agit

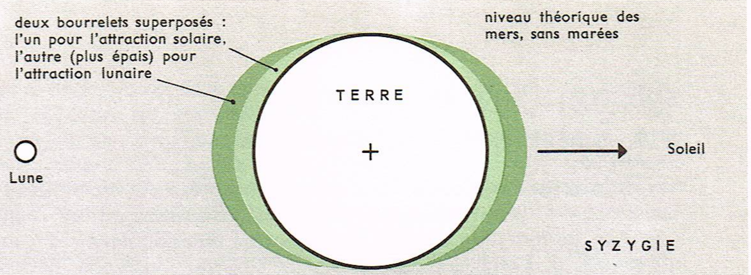
indépendamment un nombre immense de causes élémentaires) augmente indéfiniment, c'est-à-dire lorsque les fréquences se « resserrent » les unes contre les autres, le spectre s'achemine vers un ensemble irrésoluble, la courbe qui joint les segments de tout à l'heure devient continue **(spectre continu).**

horloges astronomiques et marées

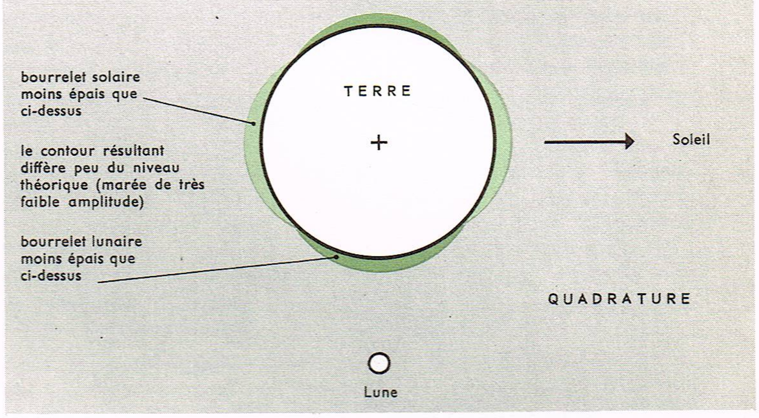
Le visiteur de l'horloge astronomique de la cathédrale Saint-Jean à Besançon est toujours surpris par les petits hublots à l'intérieur desquels il voit monter l'océan dans différents ports disséminés à travers le monde. Il ne saisit aucune relation immédiate entre les indications des hauteurs de la marée en des points aussi lointains les uns des autres. Cependant, s'il pouvait prendre suffisamment de recul, il aurait la sensation de quelque chose qui lie entre elles toutes ces indications. Ce quelque chose, il en a entendu parler dès l'école primaire : c'est l'histoire des marées fortes et des marées faibles, l'histoire de l'attraction de la Lune et du Soleil sur notre globe. Bien sûr, les astronomes savent repérer exactement l'époque des conjonctions Lune-Soleil, c'est-à-dire de leurs alignements (syzygie), qui correspondent aux marées importantes, et celle des quadratures, qui déterminent les marées faibles.

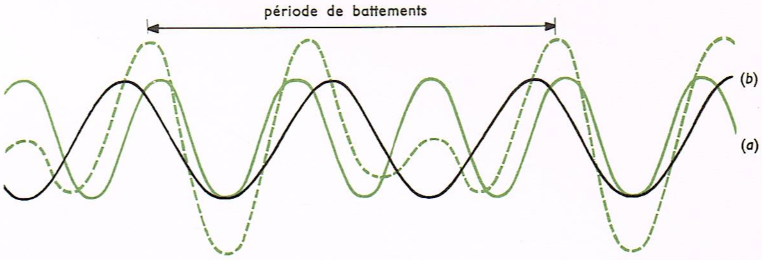
Le phénomène est périodique : en France, sur la côte océane, on a l'habitude d'attendre la mer étale deux fois par jour (fréquence fondamentale).

Pour revenir aux petits hublots qui ceinturent l'horloge astronomique, s'il est passionnant de chercher à deviner les mouvements de la mer au Havre ou à Sainte-Hélène, il n'en demeure pas moins que le visiteur attentif qui aurait la patience d'observer durant des mois et des mois se rendrait compte que les montées et les descentes d'un niveau à l'entrée du port ne sont pas aussi régulières que le suggérerait. notre figure. En un mot, tous ces mouvements qui se combinent présentent, plus ou moins cachées, plusieurs périodicités. Les connaître parfaitement permettrait de prévoir — moyennant un appui sérieux de la météorologie — tout éventuel péril maritime. Or, les fluctuations des marées, aux causes souvent multiples, ne se superposent pas toujours de manière évidente. Deux causes isolées, parfaitement définies, de fréquences légèrement différentes, donneraient lieu à un phénomène de battement. Mais, plus couramment, il s'agit de mouvements complexes, périodiques dans le cas le plus favorable. Comme dans l'affaire des soles de l'Adriatique, une analyse harmonique s'impose alors.



Tout se passe comme si la Terre était une orange mûre attirée surtout par la Lune. La peau s'étire dans le sens de la force d'attraction (et, à l'antipode, dans le sens opposé, cela étant dû au fait que la Terre et la Lune gravitent autour d'un centre commun). La marée est d'amplitude maximale dans ces régions, tandis qu'elle est « basse » dans les régions dont l'azimut est de 90° par rapport aux précédentes.

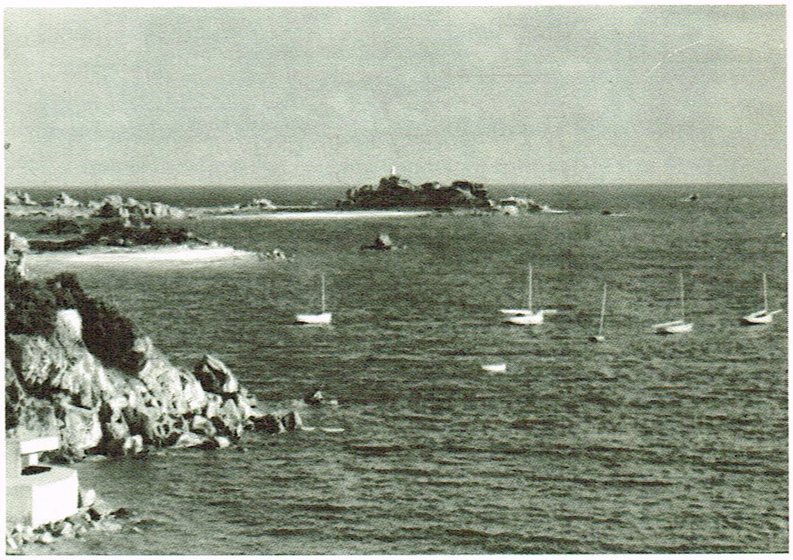




Battements de deux sources de même amplitude : $\frac{N'}{N} = \frac{3}{2}$

Mer « basse » et pleine mer dans la baie de Port-Blanc (Côtes-du-Nord).





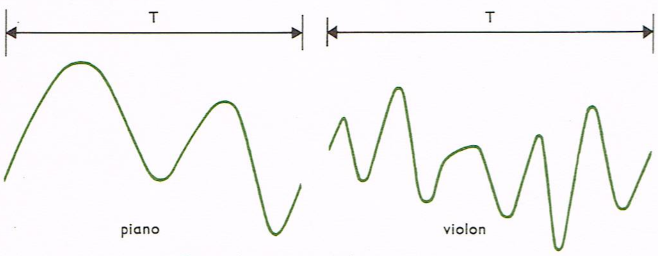
Phot. Jidébé.

spectre : son et timbre des instruments de musique

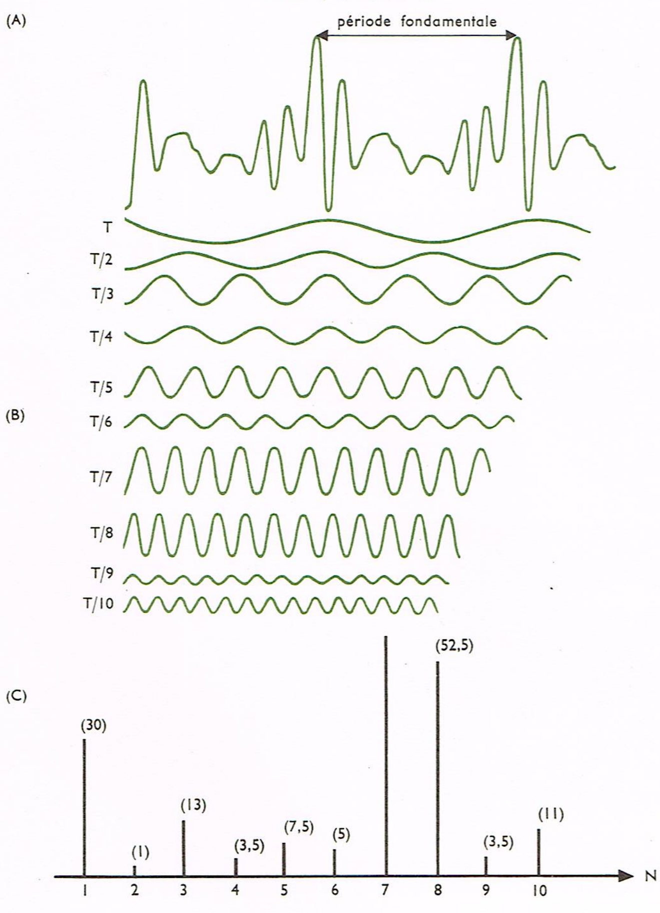
Un dernier exemple achèvera la mise en place de la méthode. La corde du violon vibre sous l'archet; celle du piano est frappée par le marteau feutré. L'anche met brutalement l'air en vibration à l'intérieur du tuyau d'orgue... Lorsque divers instruments d'un orchestre attaquent la même note, une culture musicale éprouvée n'est pas nécessaire pour distinguer l'orgue du violon ou le piano de la trompette. Ce qui les distingue, c'est le « timbre ». On dira que les différences de timbre sont dues à la présence des harmoniques, superposés au son fondamental.

Représentons l'enregistrement, sur une période, d'une note émise par le piano; le violon semble déjà plus riche en harmoniques; quant au hautbois, la série des sinusoïdes *(b)* montre l'importance relative des différents constituants de cette note. Leur superposition reconstituerait la note jouée si la synthèse respectait les proportions indiquées par le spectre.

Ici encore, l'analyse harmonique a consisté à déterminer les constituants et leur dosage, à découper en fonctions élémentaires, sinusoïdales, la fonction représentant globalement le phénomène. On parle encore d'« analyse de Fourier », du nom d'un mathématicien français qui, vers 1800, a étudié la superposition des mouvements vibratoires.



Enregistrement des sons : la même note jouée par le piano (à gauche) et par le violon. *Ci-dessous,* analyse du timbre du hautbois : (A) enregistrement; (B) différents constituants de la note; (C) spectre.



un théorème fondamental

L'importance des techniques de **Fourier** dans la physique des vibrations est telle qu'un énoncé général (1) trouve ici sa place :

Un mouvement quelconque, périodique ou non, peut toujours être considéré comme la superposition de mouvements élémentaires périodiques.

En acoustique, la note jouée par le hautbois représente un mouvement vibratoire dont la période fondamentale est T. A ce fondamental s'ajoutent des harmoniques de périodes T/2, T/3,... l'ensemble réalisant une synthèse de la note.

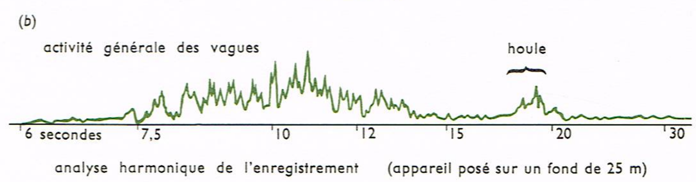
L'énoncé général précédent permet d'appliquer l'analyse de Fourier et d'adapter son expression mathématique à un effet quelconque, tel qu'une percussion ou un bruit. La percussion, par exemple, pourrait encore être « développée » en un certain nombre de vibrations simples de fréquences déterminées.

spectre et tempêtes

L'étude des vagues de l'océan a fait l'objet d'applications fort intéressantes de l'analyse harmonique. Examinons comment varie au cours du temps l'amplitude des vagues enregistrées par un navire spécialement équipé à cet effet (il reviendrait au même de prendre une photographie en coupe verticale, un profil, de la surface de l'eau; mais ce n'est guère réalisable)



Enregistrement du mouvement de l'eau à la surface de la mer et spectre correspondant. (D'après G. E. R. Deacon.)



L'analyse de Fourier a permis de chiffrer les périodes des différentes composantes du mouvement et d'en faire ressortir les amplitudes. C'est le graphique *b*, sur lequel on remarque la présence de toutes les périodes possibles comprises entre 7,5 secondes et 13 secondes environ; plus loin, autour de 19 secondes, prend place une autre bande de périodes (ou encore de fréquences), celle de la houle (notons que la période d'une « houle longue » varie de une à plusieurs minutes). Le

(1) Moyennant quelques précautions d'ordre théorique.

L'analyse d'un mouvement périodique de durée illimitée, fonction du temps, $f(t)$ par exemple (c'est pratiquement le cas d'une note émise par un instrument au cours d'un intervalle de temps durant un grand nombre de périodes du son fondamental), se traduit mathématiquement par ce qu'on appelle un « développement en série de Fourier », de la forme

$f(t) = a_0 + a_1 \sin \omega t + b_1 \cos \omega t + a_2 \sin 2\omega t + b_2 \cos 2\omega t + a_3 \sin 3\omega t + b_3 \cos 3\omega t + \ldots$, où $\omega = 2\pi/T$ est la pulsation du mouvement de période fondamentale T. L'ensemble des coefficients $a_0, a_1, b_1, a_2, b_2, \ldots$ des mouvements élémentaires (tel $a_2 \sin 2\omega t$, par exemple, qui est sinusoïdal et de pulsation 2ω, c'est-à-dire de période T/2) constitue le spectre de $f(t)$. Ce spectre est discontinu.

Si le mouvement résultait d'un nombre aussi grand que l'on veut de mouvements élémentaires, de périodes fondamentales différentes, mais aussi voisines que l'on veut les unes des autres, le spectre tendrait vers un continuum. L'expression mathématique correspondante serait un développement en intégrale de Fourier.

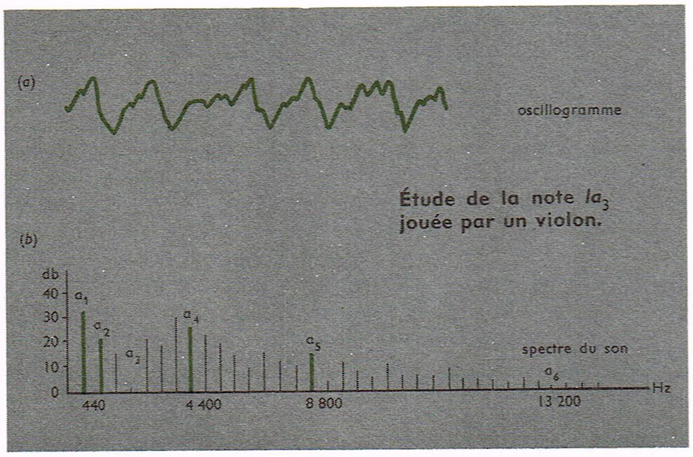
spectre est ici continu, et non plus constitué, comme dans l'analyse des exemples précédents, de valeurs discrètes.

L'intérêt des enregistrements des régimes des vagues — et surtout de leur classification — réside dans les nombreuses déductions obtenues assez facilement. Il suffit d'étudier les différences entre les temps d'arrivée des trains de vagues longues et courtes, pour évaluer la distance d'une tempête encore lointaine et la localiser : n'a-t-on pas pu mettre en évidence la naissance d'une houle en Nouvelle-Zélande en examinant des enregistrements effectués sur les côtes californiennes. Les mesures sont suffisamment rapides (l'enregistrement des profils des vagues et l'analyse durent un quart d'heure environ, et ce grâce à l'utilisation de puissantes machines électroniques, qui effectuent les calculs nécessaires, toujours très difficiles) et précises (des fluctuations de l'ordre du centimètre sont observables) pour que cette technique continue à être utilisée pour la prédiction des courants côtiers, des mouvements de la mer (elle le fut au moment du débarquement de Normandie en 1944) et des fonds marins. Associée à des méthodes statistiques, elle rend aussi les plus grands services dans l'étude des problèmes de carénage et de construction des navires.

Pour sacrifier quelque peu à l'histoire, il n'est certainement pas sans intérêt de rappeler que, dès la fin du siècle dernier, Michelson avait étudié les marées et mis en œuvre l'analyse harmonique : son maréographe aurait été construit pour enregistrer les composantes du mouvement de l'océan. C'est d'ailleurs au cours de ces expériences qu'il montra que, loin de tout continent, l'amplitude des marées était due pour 69 p. 100 environ à la déformation de la nappe liquide, et pour 31 p. 100 à la déformation propre du globe terrestre, qui « respirerait », souple comme une boule plastique, au rythme des marées.

spectre et haute fidélité

La notion de spectre est non moins familière aux acharnés de la haute fidélité, pour qui bandes passantes et décibels n'ont



Étude de la note la_3 jouée par un violon.

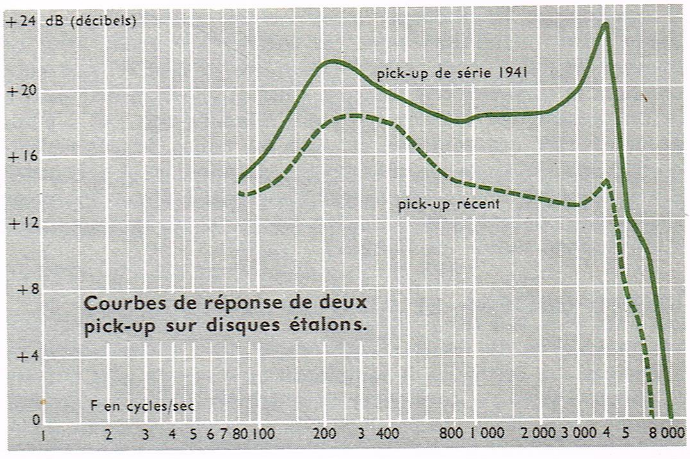
plus aucun secret. Examinons l'enregistrement oscillographique de l'intensité sonore émise par un violon jouant la note la_3 (440 vibrations par seconde) et le spectre correspondant. On y retrouve le son fondamental et un certain nombre d'harmoniques. D'autres fréquences ont également pris place dans ce spectre. Supposons alors que le son ait été transposé, transcrit sur un disque suivant une technique appropriée, puis que le disque soit « passé » sur un électrophone. Chacun des éléments de la chaîne de transmetteurs allant du violon à l'oreille — qu'il s'agisse des sillons du disque, de la tête de pick-up, des amplificateurs ou du haut-parleur — agira pour son propre compte sur notre la_3, laissant passer une certaine fraction de l'harmonique 1, une autre fraction de l'harmonique 2, etc. Finalement, l'effet en sera que les proportions des composantes du la_3 écouté différeront de celles du la_3 émis. Toute la chaîne des transmetteurs dans son ensemble, ou encore chacun d'eux séparément, peut être caractérisée par une courbe : la courbe de réponse du système. Cette courbe donne le facteur de transmission, fréquence par fréquence, c'est-à-dire, pour toute fréquence ν, le rapport entre l'importance de cette fréquence, mesurée par l'amplitude a'_ν de la vibration à la sortie du système, et l'amplitude a_ν fournie à l'entrée.

Si donc toutes les fréquences, fondamentales et harmoniques, ne sont pas transmises de la même façon, c'est-à-dire si l'amortissement est différent pour les différents sons élémentaires composant le son la_3 complexe, on ne récupérera pas à la sortie un mouvement vibratoire produisant l'impression de l'original. Il y aura distorsion.

spectre et « triage » acoustique

En acoustique, le domaine des fréquences perçues par une oreille normale est compris entre 20 Hz et 16 000 Hz environ. Il semblerait donc que le transmetteur idéal dût présenter une courbe de réponse plate entre ces fréquences extrêmes. Ce serait ne pas tenir compte de l'influence des harmoniques — y compris ceux qui, individuellement, ne seraient pas audibles —, qui donnent le caractère, le relief à la musique (nous avons déjà parlé du timbre). Mais les éditeurs de disques savent bien que, même si certaines régions du spectre subissent un amortissement, l'audition peut rester aussi agréable. On s'arrange parfois pour filtrer tel ou tel harmonique, telle bande de fréquence désagréable (certains harmoniques impairs du violon, vers le onzième, sont avantageusement éliminés).

Ainsi, la haute fidélité est finalement une judicieuse « cuisine » qui découpe l'original en harmoniques (analyse), effectue un tri (filtrage) et restitue (synthèse) une qualité de musique, de sons, susceptible de satisfaire l'oreille la plus fine, la plus subtile en matière de tonalité et d'impression rendue. Une définition de la haute fidélité? Celle de G. Slot paraît à la fois ambitieuse et conforme à ce que chacun ressentira : « ... une reproduction



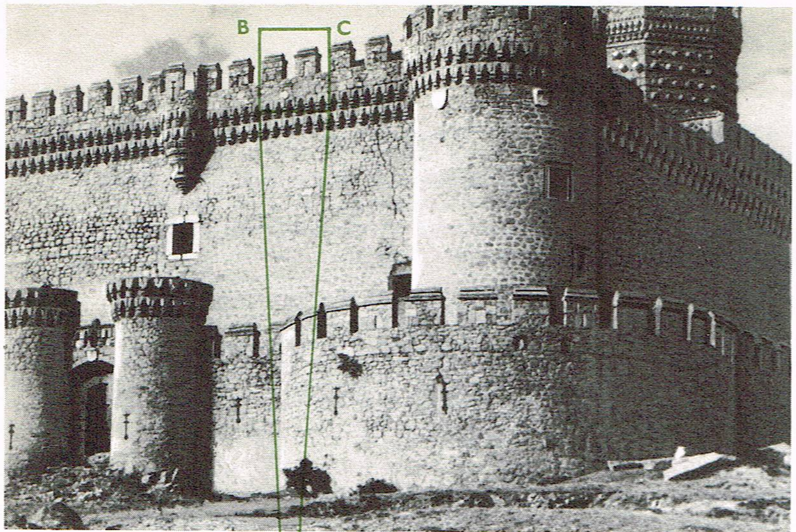
Courbes de réponse de deux pick-up sur disques étalons.

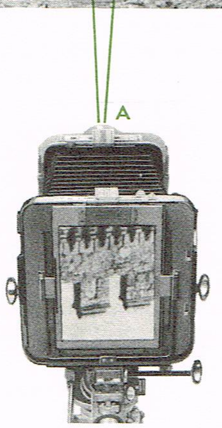
telle que la musique me procure autant de satisfaction que dans la salle de concert » (condition difficilement réalisable, puisque rien n'a été « filtré » lors de l'audition directe du concert).

spectre et couleurs

Spectre électromagnétique (ensemble de toutes les fréquences), spectre continu d'émission du corps noir, spectre discontinu de la lampe à mercure (1). Le terme de « spectre » qui est revenu si souvent au long de ces dernières pages a-t-il évoqué au moins le contenu des chapitres précédents?

(1) A l'instar des cordes vibrantes et des instruments acoustiques, et pour les mêmes raisons (caractère discret des valeurs propres), un ensemble de molécules ou d'atomes soumis à une décharge électrique ne peut émettre que sur certaines fréquences, en nombre réduit; ce fait fondamental sera revu en détail.





▲ Photographie d'un objet représentant une périodicité et, *ci-contre*, représentation de l'éclairement du sujet suivant la ligne B C. ▶

Il n'y a guère de raison de chercher dans les mots utilisés en physique un contenu qui dépasse nécessairement celui du langage courant. **Le savant qui tente d'exprimer clairement une hypothèse ou un résultat n'éprouve guère le besoin d'une langue hermétique supprimant automatiquement toute communication avec les autres hommes.** Ainsi on constate régulièrement, avec étonnement, il faut en convenir, combien le « spectre des couleurs » est connu du grand public. (« Le spectre des couleurs? Ah oui, l'arc-en-ciel... ») L'étude de ce spectre des couleurs, ou tout au moins sa description, a déjà été amorcée précédemment. Rappelons que chacune des fréquences qui le constituent correspond à une vibration électromagnétique particulière (en l'occurrence, le mouvement d'un électron autour de l'atome).

image optique et analyse harmonique

Les techniques de Fourier embrassent donc toutes sortes de vibrations. Là même où il semblerait qu'il n'y ait plus matière à vibrer, il suffit de considérer de plus près la courbe du phénomène en jeu pour constater qu'elle est la superposition d'éléments périodiques simples (principe de l'analyse harmonique).

Ainsi, regardons ce photographe opérant au pied de la haute tour crénelée.

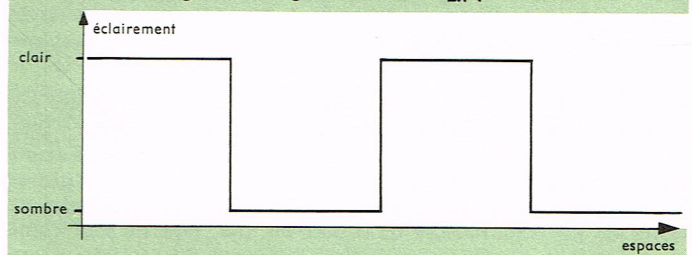
Quel intérêt scientifique peuvent bien présenter les créneaux d'une tour, ou le redan qui se profile sur le ciel clair? C'est que le photographe, lorsqu'il s'est offert l'objectif de son appareil, a décidé d'en exiger une foule de qualités; il désire surtout de bonnes images, bien nettes, « piquées », avec des quantités de tout petits détails précis. Sauf s'il fait du « flou artistique », il veut pouvoir compter, sur la photographie, le nombre de rangées de pierres qu'il a fallu superposer pour construire la tour; il veut que sur l'image le bord des créneaux ne soit pas arrondi, qu'aucun flou ne vienne altérer l'altier profil du redan.

Or, quels sont les éléments intervenant entre le bord du créneau, les pierres régulièrement ajustées — c'est-à-dire l'**objet** —, et l'**image** photographique? Dans la chaîne de transmetteurs, le plus important est certainement l'objectif photographique.

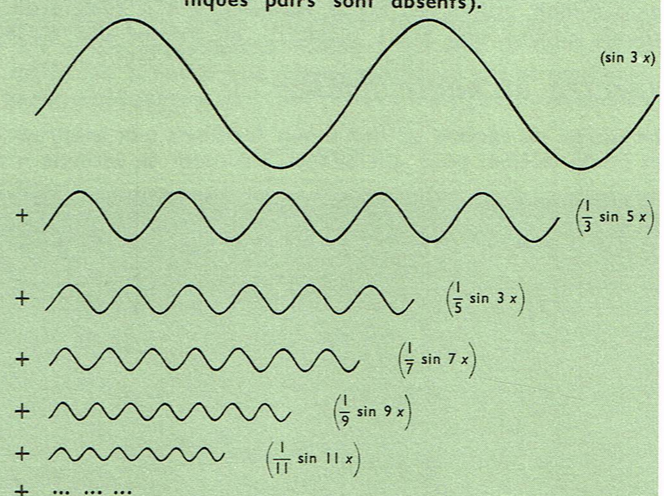
Nous avons vu qu'un transmetteur (dispositif d'enregistrement ou objectif photographique) découpe le message de l'émetteur (violon ou redan) en messages élémentaires et que tout se passe comme s'il acheminait chacun d'eux séparément vers le récepteur (oreille ou plaque photographique). Le transmetteur analyse le message reçu, agit sur les termes de fréquence différente (harmoniques ou vibrations indépendantes) suivant une « fonction de transfert », représentée par la courbe de réponse, puis, dans un dernier temps, reconstitue un message (son ou photographie), que l'on souhaite aussi fidèle, aussi semblable que possible à l'original.

Ce graphique représenté par la fonction mathématique

$$y = \frac{4}{\pi}\left(\sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x + \dots + \frac{1}{2n-1}\sin(2n-1)x + \dots\right)\dots$$

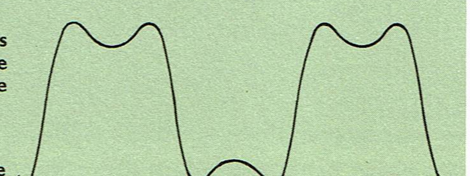


...peut être réalisé en superposant tous ceux-ci (les harmoniques pairs sont absents).



Voici les résultats lorsqu'on arrête la somme

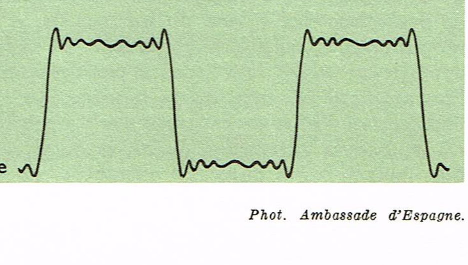
au 3e harmonique



au 7e harmonique



au 15e harmonique



Phot. Ambassade d'Espagne.

Disséquons une partie de l'objet, le haut de la tour par exemple. Suivant le plan ABC de visée, les créneaux apparaissent comme une répétition d'élément sombres alternés avec des éléments clairs. Une représentation graphique de cette observation est donnée ci-contre, où l'on a porté en abscisses les distances comptées latéralement et en ordonnées les éclairements suivant la ligne BC.

L'objectif photographique doit donc résoudre le problème suivant : transmettre ces variations périodiques d'éclairement, aux transitions brusques et nettement définies (lignes nettes aux changements de tonalité), vers la plaque (ou vers le film) photographique, et les transcrire en une « bonne image ».

Le théorème de Fourier, dont l'énoncé général a été donné plus haut, nous apprend que la représentation précédente des créneaux pouvait être encore obtenue en superposant une série de sinusoïdes de fréquences déterminées (ici, ce sont les harmoniques). Est-il besoin de nous répéter, pour préciser de nouveau le processus de la transformation? Le message à l'entrée du transmetteur, c'est-à-dire l'objet, est décomposable en messages simples, harmoniques (v. fig. p. 116). Chacun des éléments étant affecté d'un facteur de transmission, qualité propre du système, c'est-à-dire ici de l'objectif, l'image résultante sera la superposition des messages élémentaires que l'objectif aura bien voulu restituer. Ainsi, dans le cas idéal où le facteur de transmission serait le même pour toutes les fréquences, les résultats seraient ceux de la figure *b* selon que l'on ne laisserait passer que trois, sept, quinze termes. En fait, on montre théoriquement que les harmoniques d'ordre élevé sont toujours moins bien transmis, les hautes fréquences sont plus étouffées que les basses fréquences. Cela même pour un système de très bonne qualité. De plus, il existe une fréquence limite, — appelée parfois fréquence « cut-off » (ou de coupure)—, au-delà de laquelle aucune transmission n'est plus possible.

Pour se rendre compte de ce que serait le « profil en éclairement » de l'image, il suffit d'examiner de nouveau la figure. S'il manque des termes d'ordre élevé, si la transmission s'accompagne d'une amputation des hautes fréquences, les contours seront plus arrondis, plus flous. D'après la remarque de tout à l'heure, à savoir qu'il existe une fréquence limite, et compte tenu aussi du fait que la transmission n'affecte pas tous les harmoniques de la même manière, **il est interdit d'espérer restituer une image rigoureusement semblable à l'objet.** En pratique, plus l'objectif laissera passer d'harmoniques élevés, mieux seront précisés les détails et rendus nets les contours de l'objet, puisqu'on se rapprochera d'autant plus de la similitude parfaite objet-image. A la limite, un détail apparaîtra ou n'apparaîtra pas suivant que ses dimensions correspondent à une fréquence transmise ou non.

Prenons un exemple. Sur une image tirée d'un film classique d'amateur (largeur de l'image : 8 mm), les dimensions des doigts d'un personnage peuvent être de 0,02 mm. Dans les mêmes conditions habituelles de prise de vues, un poil de barbe aurait environ 1/800 de millimètre de long; encore faudrait-il qu'il fût couché, ou photographié de biais. Pour voir les doigts, et non pas seulement les deviner, il est donc nécessaire que l'objectif soit capable de « passer » des détails correspondant dans le plan de formation de l'image à une période d'espace de 0,02 mm, c'est-à-dire encore à une fréquence de 50 éléments par millimètre. Quant aux poils de barbe, même si l'objectif pouvait les « résoudre », il faudrait ensuite que le film photographique fût capable d'enregistrer un détail de cette dimension. Alors, avec les données ci-dessus, pourtant parfaitement raisonnables, il y a peu de chances que le marchand de pellicules sache refréner son sourire si vous lui demandez un film de 8 mm d'une si grande qualité.

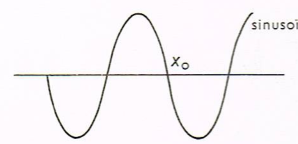
Une exploitation plus poussée du raisonnement précédent conduirait aux notions de transfert d'image et de filtrage optique. Et ces considérations dépassent le cadre de ce chapitre.

une question simple, et une réponse qui ouvre bien des portes

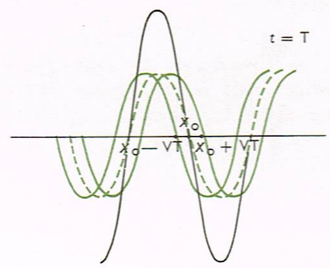
Nous savons qu'un instrument de musique ne peut émettre que des harmoniques du son fondamental; quand on demande le *la*$_3$ à un violon, il répond par les seules fréquences multiples de 440 (440 vibrations par seconde pour le fondamental; 880, 1 320, ... pour les harmoniques, dont l'amplitude décroît assez vite avec l'ordre). En revanche, le spectre ne comporte pas de terme correspondant à 430 ou à 450 vibrations.

Pourquoi donc?

la vibration initiale $[y = a \sin \frac{2\eta\nu}{V}(x - x_o)$ pour $t = 0]$



se propage à droite et à gauche et donne naissance à l'onde « somme »



$$y = a \sin \frac{2\eta\nu}{V} [(x - x_o) + Vt] + a \sin \frac{2\eta\nu}{V} [(x - x_o) - Vt]$$
$$= 2\,a \sin \frac{2\eta\nu}{V} (x - x_o) \cos 2\eta\nu t$$

La vitesse de propagation V ne dépend que de l'état physique de la corde (matière, tension, masse et diamètre) et en aucun cas de la nature de l'onde transmise, de la percussion donnée par l'archet.

Il est facile de le voir : admettons, pour simplifier que la corde pincée vibre dans le plan vertical représenté par la figure. L'ordonnée de la courbe est obtenue en faisant la somme des quantités : $y = 2a \sin \left[\frac{2\pi\nu}{V}(x\text{-}x_o)\right] \cos 2\pi\nu t$.

Mais la corde est fixée en O (x = O); y (λ) doit être nulle en ce point; d'où

$$\frac{2\pi\nu}{V}(\mathrm{O}\text{-}x_o) = \text{multiple de } \pi.$$

De même, la corde est fixée en O′ (x = L); y (λ) doit être nulle en ce point, et

$$\frac{2\pi\nu}{V}(\mathrm{L}\text{-}x_o) = \text{multiple de } \pi.$$

En soustrayant terme à terme les deux égalités précédentes, il apparaît nécessaire que $\frac{2\pi\nu\,L}{V}$ soit un multiple de π, soit Kπ $\left(\nu = K\frac{V}{2L}\right)$. L étant la longueur de la corde OO′, et V, la vitesse de propagation des ondes le long de cette corde.

Il est donc impossible que des ondes de toute autre fréquence soient émises. De plus, les fréquences possibles, en particulier la fondamentale, ne dépendent pas de l'attaque de la corde par l'archet (susceptible de faire varier le timbre, non la note), mais seulement de la tension et de la nature de la corde (dont dépend la vitesse de propagation V) et de la

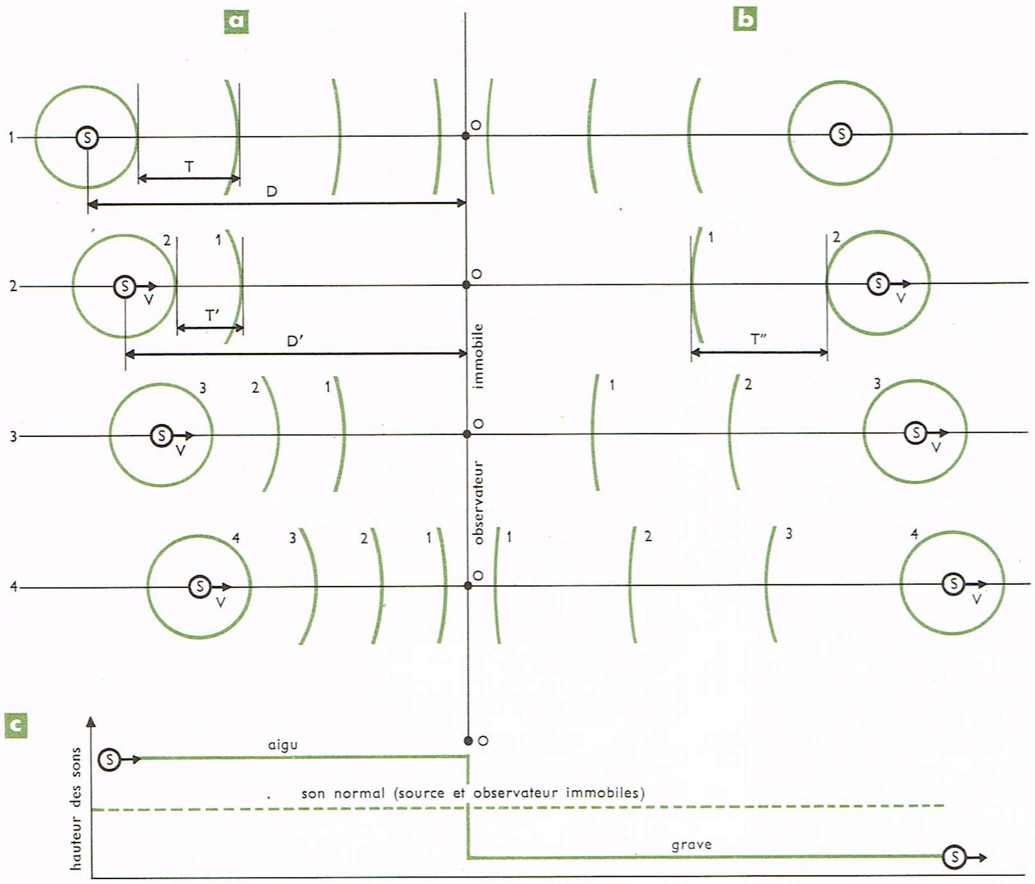
longueur OO′, c'est-à-dire de la distance à laquelle nos doigts pincent la corde.

Une théorie si raffinée n'explique-t-elle que le fonctionnement des instruments à corde? Nous nous en serions voulu de faire perdre son temps au lecteur, non nécessairement mélomane, si une telle analyse ne nous avait ouvert la porte sur des connaissances beaucoup plus fondamentales, celles qui ont trait aux **phénomènes si riches et si nombreux de résonance et aux théories de la mécanique quantique.**

Mais n'est-ce pas le propre d'une théorie que de rassembler sous une même formulation les phénomènes les plus divers, du plus naturel, du plus quotidien (violon, flûte, photographie, etc.) au plus abstrait, au plus lointain, au plus rare (théories de la mécanique quantique et théorie des résonateurs harmoniques).

Résumons donc notre récente acquisition dans une formule d'apparence abstraite, mais claire : **un système physique, telle une corde de violon, n'est pas susceptible de prendre n'importe quel état. A chaque état possible, correspond un nombre appelé « fréquence propre » ou « valeur propre ». L'ensemble des valeurs propres dépend des conditions aux limites** (conditions aux extrémités de la corde vibrante, à l'extrémité du tuyau d'orgue, à la surface du guide d'onde, ...). Les fréquences propres sont souvent multiples d'une fréquence dite « fondamentale ».

UNE CURIEUSE PROPRIÉTÉ DES PHÉNOMÈNES ONDULATOIRES : L'EFFET DOPPLER-FIZEAU



1. La source S qui produit les sons et l'observateur O sont immobiles : la fréquence et, partant, la longueur entre deux ondes, et le temps T qui les sépare sont constants en (*a*) comme en (*b*) et pour toute autre position relative de S et O.
2, 3, 4. En (*a*), la source se déplace vers O (toujours immobile); en (*b*), elle s'en éloigne. Dans le premier cas, S se rapproche de O dans l'intervalle de deux ondes successives et la distance qui sépare ces dernières en est réduite d'autant; le temps T aussi, d'où la perception par O d'un plus grand nombre d'ondes par unité de temps. La fréquence est donc plus grande que précédemment (1) et le son perçu sera plus aigu (*c*).

Réciproquement, en (*b*) la source s'éloigne de l'observateur entre l'émission de deux ondes successives : T″ est donc plus grand que T, et le nombre d'ondes captées en O par seconde est plus petit qu'en 1. Les sons perçus par l'observateur sont donc plus graves (*c*). Les phénomènes seraient identiques si la source était immobile et l'observateur en mouvement ou si S et O se déplaçaient. Dans tous les cas possibles, ce qui compte est le déplacement relatif de l'un par rapport à l'autre.

C'est un phénomène bien connu des voyageurs et des automobilistes : lorsque le train ou la voiture où ils se trouvent croisent une locomotive qui siffle, le son du sifflet leur paraît baisser brusquement au moment du croisement. Il en va de même lors du croisement de deux automobiles, le sifflet ayant alors été remplacé par l'avertisseur.

L'expérience courante montre que si une source émet un son, la hauteur du son perçu par l'observateur est modifiée par le mouvement de la source; elle l'est également par le mouvement de l'observateur. Le son est plus aigu quand l'observateur et la source s'approchent l'un de l'autre, plus grave quand ils s'éloignent.

Il s'agit là d'un phénomène très important, qui se manifeste pour tous les phénomènes vibratoires : l'effet Doppler-Fizeau. Nous le retrouverons en optique, car il existe pour la lumière aussi bien que pour les vibrations sonores. Ses applications astronomiques sont très nombreuses.

Il suffit, pour se l'expliquer, de se représenter un son musical comme une série d'ébranlements brefs de l'air, identiques entre eux, émis par la source à des intervalles de temps égaux. L'intervalle de temps qui sépare deux ébranlements successifs est la période T du phénomène vibratoire. Lorsque la source et l'observateur sont immobiles, l'observateur reçoit ces ébranlements à des intervalles de temps égaux à ceux de l'émission T. Il n'en est plus de même si la source, ou l'observateur, sont en mouvement, et nous allons montrer pourquoi.

Par rapport à un observateur immobile au point O, une source sonore, située à la distance D de l'observateur, arrive sur lui à la vitesse v. A cet instant, la source émet un signal, une perturbation qui se propage à la vitesse c et arrive à l'observateur au bout du temps $t_1 = \frac{D}{c}$. Au bout d'un temps égal à la période T, la source émet une deuxième perturbation qui se dirige de la même façon vers l'observateur. Mais elle n'a plus la distance D à franchir, car, pendant l'intervalle de temps qui sépare l'émission des deux signaux, la source s'est déplacée de la distance $SS' = vT$.

Pour parcourir la distance $D' = D - SS' = D - vT$, il faudra à la deuxième onde un temps $t'_2 = \frac{D'}{c} = \frac{D - vT}{c}$. Pour l'observateur, la période apparente du phénomène vibratoire n'est autre que le temps qui sépare l'arrivée des deux ondes. La première arrive au bout du temps t_1. La deuxième, émise par la source une période plus tard, a besoin du temps t'_2 pour faire le chemin. Elle arrive donc finalement à l'observateur à l'instant $t_2 = T + t'_2 = T + \frac{D - vt}{c}$. Et la période apparente du phénomène vibratoire sera, pour l'observateur,

$$T' = t_2 - t_1 = T + \frac{D - vT}{c} - \frac{D}{c} = T - \frac{vT}{c}.$$

Première constatation : le phénomène ne dépend que de la vitesse de la source et nullement de la distance de celle-ci à l'observateur.

Deuxième constatation : la période T' ne diffère de T que par le terme $-\frac{vT}{c}$. **Seul intervient donc le rapport de la vitesse de la source à la vitesse de propagation de l'onde.**

Si nous avons affaire à des ébranlements sonores, qui se propagent dans l'air à une vitesse de l'ordre de 300 m/s, nous trouverons très facilement des sources dont la vitesse v soit une fraction notable de c. Si la source se déplace à 100 km/h, par exemple, c'est-à-dire à 28 m/s, le terme $\frac{v}{c}$ vaut sensiblement 0,1, et la période est amputée du dixième de la valeur qu'elle aurait en l'absence du mouvement. La fréquence (inverse de la période) est augmentée dans des proportions semblables, et l'effet acoustique est très sensible : le son paraît plus aigu.

Revenons alors au croisement de nos deux automobiles. Pendant toute la période qui précède le croisement, la source s'approche de l'observateur. Après le croisement, elle s'en éloigne, et nous devons compter la vitesse v négativement, donc mettre le signe + dans la formule donnant T', au lieu du signe —. La période paraît augmenter après le croisement, et la hauteur du son semble baisser. C'est donc un saut brusque, nous pourrions dire presque instantané de la hauteur du son, qui se produit à l'instant du croisement : ce n'est qu'à l'instant précis du croisement, pendant une infime fraction de seconde, que l'on pourrait entendre le son à sa véritable hauteur.

Considérons maintenant des ébranlements lumineux, et non plus sonores; la vitesse de propagation de ces ondes est considérablement plus grande; elle atteint la valeur de 300 000 km/s. En comparaison de cette valeur énorme, la vitesse de nos pauvres mobiles, terrestres ou aériens, est tout à fait infime. Nous ne constaterons jamais de variation de fréquence lumineuse à la suite du mouvement de tels véhicules. Et si nous pouvions la constater, que verrions-nous? Le lecteur, s'il a parcouru les premières pages du chapitre consacré aux ondes électromagnétiques, nous répondra sans hésiter : un changement de couleur! C'est en effet, pour les phénomènes optiques, la couleur qui est la manifestation sensible de la fréquence des ondes.

Il nous faut maintenant aller de l'avant et chiffrer le phénomène de Doppler-Fizeau en utilisant la terminologie de l'optique, c'est-à-dire parler en longueurs d'onde.

On passe de la période T à la longueur d'onde en multipliant la période par la vitesse c de propagation des ondes ($\lambda = cT$). Si donc une source animée de la vitesse v émet de la lumière de longueur d'onde λ, cette lumière paraîtra à un observateur immobile avoir la longueur d'onde $\lambda - vT$.

Si la distance de la source à l'observateur diminue, la longueur d'onde semble plus petite; on dit que la lumière est « décalée » vers le bleu. En optique, nous l'avons dit, nous trouvons rarement des sources dont la vitesse soit comparable, même de loin, à celle de la lumière. Cela ne se produit que dans des cas très particuliers, extrêmement importants pour l'étude de la physique. Citons essentiellement :

1° Les vitesses d'agitation des molécules ou des atomes d'un gaz porté à très haute température. Ces atomes sont des émetteurs naturels de lumière; ils émettent selon certaines longueurs d'onde bien précises, qui dépendent de leur nature chimique. Leur mouvement d'agitation désordonné provoque un étalement de ces longueurs d'onde de part et d'autre de la longueur d'onde qui serait émise s'ils étaient au repos;

2° Les vitesses que l'on rencontre en astronomie et qui se chiffrent par milliers de kilomètres à l'heure. Il sera longuement parlé de l'effet Doppler-Fizeau en astronomie :

$$\lambda' = cT' = cT\left(1 - \frac{v}{c}\right).$$

La variation relative de longueur d'onde ($\Delta\lambda/\lambda$) est l'opposé de la vitesse relative (v/c).

Remarquons que seul intervient le mouvement relatif de la source et de l'observateur l'un par rapport à l'autre : nous savons que la distance entre la Terre et une étoile s'accroît ou diminue mais rien ne nous permet de dire que l'étoile s'approche de nous, ou qu'au contraire c'est nous qui allons vers elle. Il en est de même au cours du croisement de deux automobiles : si toutes deux vont l'une vers l'autre à 100 km/h, la variation apparente de la hauteur du son sera le double de ce qu'elle serait pour un automobiliste arrêté au bord de la route.

Un automobiliste voit un feu rouge l'arrêter. Peut-il accélérer suffisamment pour pouvoir prétendre que le feu est vert et l'autorise à passer? Certes non.

1. L'agent, immobile sur la chaussée, continuera à voir le feu rouge et dressera une contravention pour « barrage forcé ».

2. Plus une seconde contravention pour excès de vitesse (λ rouge = 0,6; λ vert = 0,54 $\frac{\delta\lambda}{\lambda}$ où $\frac{\delta\lambda}{\lambda} = 0,1$ correspond à une vitesse de 30 000 km/s). En admettant le moteur assez puissant pour que la voiture puisse, en quelques mètres, acquérir une telle vitesse!

Une des applications les plus récentes et les plus intéressantes de l'effet Doppler-Fizeau est la mesure de la hauteur et de la vitesse des satellites artificiels. Cette dernière est faible par rapport à celle de la lumière, mais l'effet produit est néanmoins aisément mesurable. (Ce n'est pas une source lumineuse qu'on utilise effectivement dans l'incapacité où l'on est encore de canaliser la lumière sur les distances de l'ordre de 300 à 500 km qui séparent les satellites de nous. Avec le développement récent des lasers, cela deviendra probablement possible avant peu. On utilise tout simplement, placé sur le satellite, un émetteur radio d'ondes électromagnétiques, ondes qui se propagent avec la même vitesse que la lumière et qui sont de nature identique.) Comment mesure-t-on la variation de longueur d'onde? Les radio-amateurs le devinent aisément : par battement avec un oscillateur local au sol.

Nous avons admis jusqu'à présent que le mouvement de la source se produisait dans la direction de l'observateur. Mais, dans le cas du satellite, il n'y a pas de croisement à proprement parler. Il faut alors appliquer à l'effet Doppler-Fizeau une formule un peu plus complexe, où seule intervient, dans la variation de la fréquence, la composante de la vitesse dans la direction de l'observateur. En astronomie, on appelle cette composante la « vitesse radiale ». Les formules sont plus compliquées, mais aussi plus fécondes, et, comme nous l'avons dit, l'étude minutieuse de la fréquence apparente reçue de l'émetteur radio permet de déterminer à la fois la hauteur et la vitesse du satellite.

●

Ainsi, l'effet Doppler-Fizeau se présente-t-il comme une propriété commune à tous les phénomènes périodiques : ondes acoustiques, lumière, ondes radio. Son champ d'application est extrêmement vaste, et le lecteur qui aura bien voulu lire ce chapitre un peu austère s'en trouvera récompensé lorsque, dans les pages qui suivent, il sera fait allusion à cet important phénomène.

LES RÉSONANCES

En 1852, à Angers, une troupe de soldats aborde un pont au pas cadencé. **Très rapidement, le pont entre en vibration et s'écroule.**

De même, un enfant suffisamment patient qui ferait tomber de l'eau goutte à goutte, à une certaine cadence, en un point judicieusement choisi au milieu d'un rail, verrait celui-ci vibrer de plus en plus fort et finir par se briser.

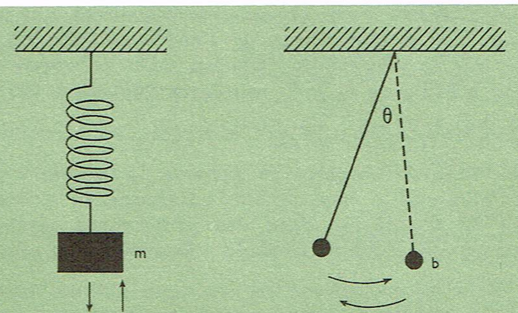
Sans pousser avec une grande force une balançoire, on arrive à la faire bouger un peu. Lorsque l'impulsion, aussi faible soit-elle, se répète « à de bons instants », c'est-à-dire chaque fois que la balançoire repart dans le sens de l'impulsion, l'amplitude du mouvement augmente. Jusqu'à un certain moment seulement; l'apport de chaque nouvel élan semble alors compenser tout juste les causes de ralentissement.

Quels sont ces phénomènes? Des résonances.

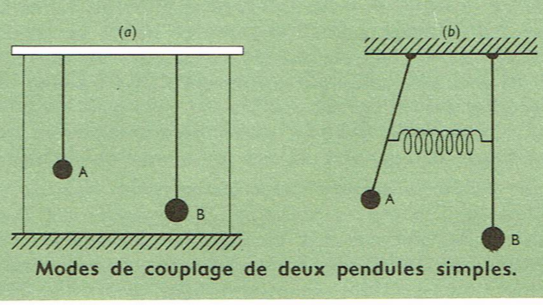
Il serait présomptueux de vouloir donner en quelques pages plus qu'une idée de l'un des phénomènes les plus importants de la physique, aux conséquences aussi nombreuses que diverses. Qu'il faille traduire en variations de pressions de l'air (c'est-à-dire finalement en ondes sonores) les variations de tension électrique (signal radioélectrique) aux bobinages d'un petit aimant, ou de deux côtés d'une accumulation de petits grains de charbon, ou qu'il s'agisse de rendre agréable à l'oreille le grincement d'une corde sous l'archet, le filet d'air sortant de la gorge du ténor, des phénomènes de résonance interviennent — résonance de la membrane et du pavillon du haut-parleur, de la boîte finement travaillée du violon ou de l'organe vocal bien entraîné du chanteur.

Est-il besoin de multiplier les exemples avant de tenter de se représenter le fait élémentaire, physique, qui sait si bien se mettre au service de l'homme ou bien le conduire à la catastrophe?

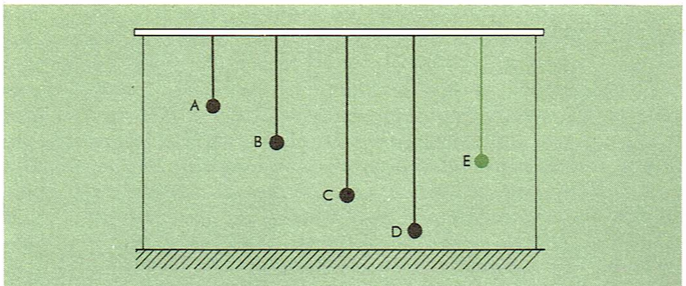
Imaginons une masse suspendue au bout d'un ressort, ou encore une bille au bout d'un fil, un pendule. La masse ou la bille, écartées de leurs positions initiales, vont tenter d'y revenir. Un système d'oscillations s'établit, puis s'amortit et finit par mourir au bout d'un laps de temps plus ou moins long. On dit qu'on a affaire à un **système à un seul degré de liberté,** exprimant ainsi que la seule donnée de l'angle O suffit à définir la position de la bille, astreinte à se déplacer sur un cercle vertical et fixe. Tout autre est le problème d'un système à deux degrés de liberté. Une représentation simple en est donnée par deux pendules associés. On dit encore qu'on réalise ainsi un couplage. Les pendules A et B sont tous deux attachés à une barre maintenue par deux fourchettes. Lorsqu'on lance le pendule A sans toucher à B, ce dernier commence par osciller faiblement, puis son amplitude croît jusqu'à une valeur maximale, tandis que décroît l'amplitude du premier, comme si A abandonnait son énergie au profit de B. Ensuite, l'amplitude de B diminue et simultanément celle de A augmente. L'effet s'inverserait un grand nombre de fois s'il ne s'amortissait assez rapidement. Ainsi les deux pendules réagissent l'un sur l'autre par l'intermédiaire de la barre qui les supporte tous deux. Le premier a joué le rôle de moteur, d'excitateur. Le second s'est comporté en résonateur.



Exemples de pendules simples : masse suspendue à un ressort oscillant librement et boule pesante au bout d'un fil.



Modes de couplage de deux pendules simples.



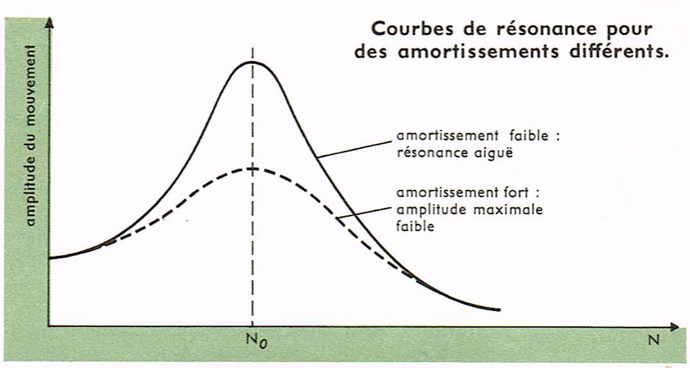
Les résonateurs A, B, C, D ne réagissent pas exactement de la même façon lorsqu'on fait varier la longueur du pendule excitateur E.

L'expérience serait encore plus spectaculaire si l'on utilisait une série de tels résonateurs constituée par des pendules de longueurs différentes : A, B, C, D, ... (système à plusieurs degrés de liberté). Lorsqu'on fait varier la longueur de l'excitateur E, les pendules A, B, C, D, ... entrent en mouvement à leur tour; le maximum d'efficacité, c'est-à-dire encore d'amplitude, pour chacun d'eux se produit pour celui qui a la même longueur que E. Or, la fréquence de battement d'un pendule est fonction de sa longueur (1); elle est donc bien déterminée pour un pendule donné qui oscillerait librement. On parlera alors de **fréquence propre** (caractéristique). Le système réagira d'autant plus facilement que la cadence de l'excitateur sera voisine de cette fréquence propre. Lorsque les fréquences de l'excitateur et du système sont exactement les mêmes, le phénomène est maximal : on se trouve dans les conditions de résonance. Ces conditions correspondent à une fréquence propre.

Dans le cas de deux pendules couplés, il n'y a plus aucune raison pour que les fréquences correspondant aux maximums d'amplitude de A et de B soient les fréquences propres de A et de B considérés séparément. On peut démontrer que le couplage ne laisse pas subsister les fréquences propres. Lorsque le nombre d'éléments — en l'occurrence de pendules — est supérieur à deux, le mouvement des pendules devient de plus en plus complexe. Restons donc un instant encore dans une situation relativement simple.

Lorsque tout à l'heure on lançait la balançoire, « répéter les impulsions au bon moment » signifiait qu'il fallait ajuster la fréquence excitatrice, celle qui correspondait aux oscillations forcées, de telle façon qu'elle fût voisine de la fréquence propre. Alors l'amplitude résultante augmentait jusqu'à la valeur de résonance.

Cela peut être représenté sur un graphique suivant les axes duquel on a porté en abscisses les fréquences N (N_0 étant la



Courbes de résonance pour des amortissements différents.

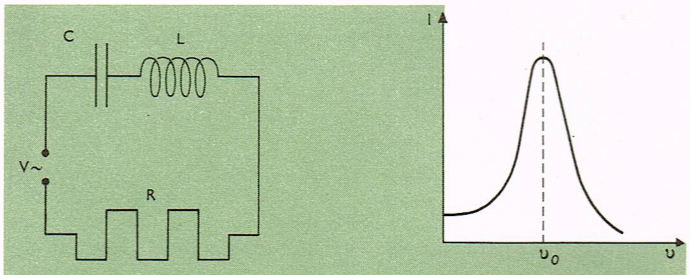
(1) Pour un pendule simple constitué par une petite sphère pesante suspendue au bout d'un fil de longueur l, la période T, inverse de la fréquence $\left(T = \frac{1}{N}\right)$, est donnée par $T = 2\pi\sqrt{\frac{l}{g}}$, où g est l'accélération due au champ de pesanteur.

fréquence de résonance) et, en ordonnées, les amplitudes du mouvement. L'amplitude maximale est d'autant plus grande que les résistances au mouvement, les amortissements, sont faibles. Pour $N = N_0$, les causes d'affaiblissement sont juste compensées par la cause excitatrice qui apporte de l'énergie fraîche.

Les troupiers du pont d'Angers avaient malencontreusement, et fortuitement accordé leurs pas à la fréquence de résonance de l'ensemble. Et la solidité de la construction n'avait pas résisté à l'amplitude de l'oscillation résultante.

Se protéger de ces résonances mécaniques aux imprévisibles conséquences n'est parfois qu'une affaire d'équilibrage (arbres de machines tournant à grande vitesse, par exemple); il est, en revanche, des systèmes qui se prêtent moins bien à leur suppression totale : les vibrations de la portière ou du coffre de voiture qui font grand bruit à une vitesse assez bien déterminée, le frémissement (heureusement vite amorti) de la clayette du réfrigérateur lorsque démarre ou s'arrête le moteur. Mais cela n'est pas très dangereux.

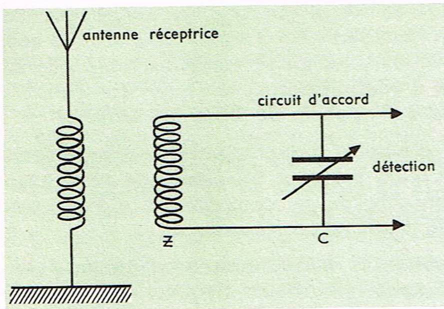
Dans le domaine des oscillations électriques, les phénomènes de résonance peuvent être décrits d'une manière tout à fait semblable. Si l'on applique une tension alternative de fréquence ν aux bornes d'un circuit comportant différentes impédances, on peut s'attendre à voir s'établir un mouvement oscillatoire.



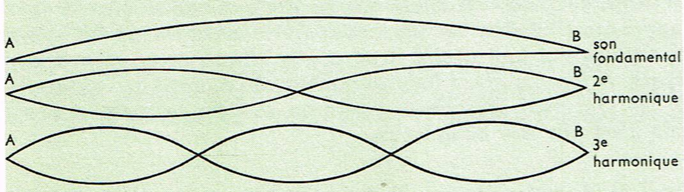
Circuit électrique oscillant et sa courbe de résonance.

Mais le circuit s'oppose au passage des électrons, et c'est précisément cette « résistance » qui traduit l'impédance : avant de s'amortir, d'autant plus vite que l'impédance qu'il doit vaincre est grande, le courant a tendance à osciller, avec une fréquence ν_0, telle que l'opposition soit la plus faible possible (« loi du moindre effort »). Si donc la tension alternative appliquée a une fréquence voisine de ν_0, il suffit d'un faible apport d'énergie pour entretenir les oscillations : on a affaire à un phénomène de résonance. Tout se passe comme s'il se produisait alors, pour $\nu = \nu_0$, une multiplication de la tension, comme si une « surtension » était appliquée aux bornes de circuit; on parlera de coefficient de surtension, ou facteur de « qualité » du circuit (des coefficients analogues interviennent non seulement en électricité, mais aussi en optique, dans la résonance mécanique, etc.).

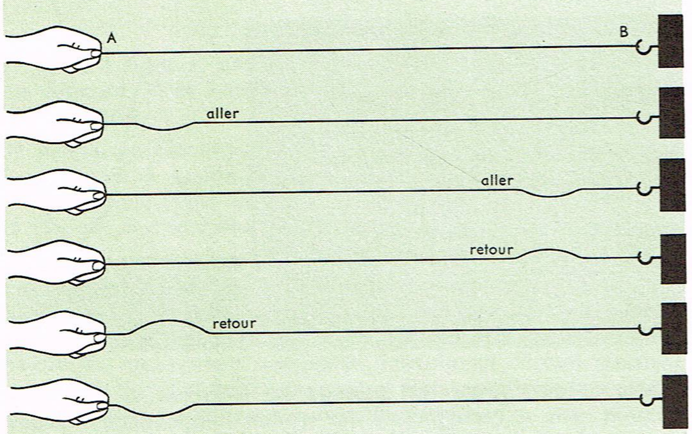
Pour un circuit donné, plus ce coefficient est grand, plus la courbe est pointue et, toutes proportions gardées, étroite; il représente donc aussi les qualités de sélectivité d'un circuit, la variation d'intensité étant très rapide pour une fluctuation faible de la fréquence autour de ν_0. On imagine immédiatement le parti que l'on peut tirer de cette propriété, par exemple en radioélectricité : supposons qu'une antenne réceptrice reçoive des signaux sur une gamme étendue de fréquences; dans le circuit d'entrée du récepteur, couplé avec l'antenne, le condensateur variable C permet de changer maintenant la fréquence de résonance ν_0, afin d'accorder l'ensemble sur une fréquence choisie à l'avance. Le signal d'intensité maximale, celui qui correspond à cette fréquence particulière, est alors envoyé vers les étages suivants, où il sera amplifié, détecté, etc., amené finalement à l'étage de sortie où se trouve le haut-parleur.



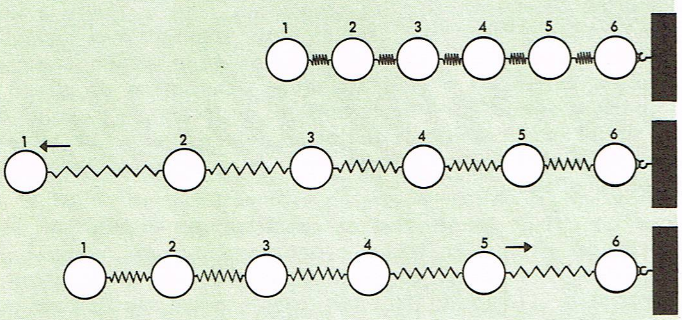
« Entrée » d'un poste récepteur de radio : l'organe essentiel est un circuit oscillant.



Période fondamentale et harmoniques. Lorsque la fréquence double, triple, etc., il se forme respectivement 2, 3, etc., fuseaux.



Réflexion d'une ondulation provoquée, sur une corde élastique, par une secousse de haut en bas et de bas en haut.



Lorsque la boule 1 est écartée de sa position d'équilibre, toutes les autres boules sont entraînées par les ressorts, mais chacune d'elles a un retard sur la boule précédente. La boule 6 étant fixe, le processus se reproduit dans le sens contraire. Les oscillations de chaque boule de part et d'autre de sa position d'équilibre ne peuvent dépasser une fréquence limite : si l'on imprime à la boule 1 un mouvement de fréquence plus grande, les autres boules ne propageront plus son mouvement, par trop rapide.

Voilà donc toute une chaîne de transmission : le studio, son ténor, le piano d'accompagnement (il se trouve d'ailleurs que la caisse de résonance du piano est telle que son action, en tant que résonateur, est pratiquement la même sur toutes les fréquences du spectre acoustique), les micros, les circuits d'accord de l'émetteur, les circuits du récepteur, les haut-parleurs, les baffles, l'oreille du mélomane. Chaque élément, chaque maillon de la chaîne est gouverné par des conditions de résonance, qu'elles interviennent sur les dimensions, les formes ou les structures. **Rares sont les mesures physiques ou les appréciations de qualité qui ne mettent à contribution les phénomènes de résonances.** Mais ne représentent-ils pas, ces phénomènes, l'inévitable, l'éternelle interaction du domaine des ondes avec la matière?

Un pendule simple : une période parfaitement déterminée de résonance; deux pendules simples : deux périodes propres différentes de la période de chacun d'eux; trois pendules simples réagissant mutuellement : trois périodes propres toujours distinctes. Et si maintenant, au lieu de ces pendules simples où la masse est concentrée en un point, nous envisageons un corps réel capable d'osciller ou de vibrer, une tige métallique, une corde de violon, un diapason, une cloche, combien cela fera-t-il de périodes propres, combien de résonances possibles? comment les exciter, comment relier les fréquences aux dimensions de l'objet, aux propriétés mécaniques et physiques du métal?

Question dont les réponses remplissent de volumineux traités qui n'épuisent pas la question. Tant s'en faut : la théorie des résonances est pleinement adulte, riche de résultats et de méthodes puissantes, riche aussi de possibilités et de promesses d'avenir. Il y a autant de travail pour les expérimentateurs que pour les théoriciens. Les problèmes posés à ces derniers pour l'objet le plus simple (un diapason, voire un disque circulaire) excèdent les moyens actuels de l'analyse mathématique. Quant aux expérimentateurs, ils apprennent chaque jour combien les modèles mathématiques les plus élaborés ne sont qu'un reflet assez lointain de la complexité du réel.

Que cette mise au point ne nous décourage pas, bien au contraire, de chercher à nous faire une idée de la nature des phénomènes, d'une part en interprétant quelques expériences simples, d'autre part en exploitant des modèles très idéalisés, mais cependant très instructifs.

Quoi de plus simple que de produire des modes de vibrations d'une corde tendue? Une drisse de bateau convient parfaitement, n'importe quelle corde souple — et pas trop légère — fait l'affaire tout aussi bien; nous la fixons à une extrémité B, nous la tendons légèrement et donnons à la main un léger déplacement rythmé à l'autre extrémité A. D'abord une petite impulsion brève. On voit l'onde solitaire se former, se réfléchir en B et revenir en A; si l'on tend davantage, la vitesse de propagation augmente. Si l'on prend une corde plus lourde (et qui reste très souple, ce point est essentiel; l'élasticité de la corde, en revanche ne fait rien à la chose), la vitesse est moindre.

Avec des cordes légères et fortement tendues, comme celles du violon, la vitesse est bien trop élevée pour que l'aller et le retour de l'onde solitaire soient visibles.

Au moment précis où l'impulsion réfléchie est revenue en A, donnons une nouvelle impulsion qui repart vers B, de telle sorte que le mouvement de la corde soit un régime permanent dans lequel il y ait toujours une impulsion en train de voyager alternativement de A en B et de B en A.

Au lieu d'une suite discontinue d'impulsions, donnons à l'extrémité A un mouvement de va et vient continu (mouvement sinusoïdal), de telle manière que sa période soit encore égale au temps de parcours aller et retour d'une impulsion. Que voyons-nous? C'est tout l'ensemble de la corde qui prend un mouvement régulier dans lequel chacun de ses points oscille avec la même période; l'amplitude est nulle en B et maximale



A l'instar de M. Jourdain avec sa prose, cette fillette fait de la résonance, sans le savoir, avec sa balançoire, lorsqu'elle produit ses efforts en accord avec la période propre de l'escarpolette.

au milieu, où elle est bien supérieure à l'amplitude du mouvement que la main de l'opérateur imprime au point A. N'est-ce pas tout simplement une expérience de résonance que nous venons de décrire? La corde ne se comporte-t-elle pas comme un pendule dont la période est fixée par la durée de propagation d'une impulsion? Conclusion hâtive, car il faut continuer l'expérience. Secouons maintenant l'extrémité A avec une fréquence exactement double de la précédente. Nous obtenons de nouveau une résonance, c'est-à-dire des amplitudes largement supérieures à celles de A, mais maintenant les maximums sont au quart et aux trois quarts de la longueur, le milieu restant pratiquement immobile. Ainsi la corde est aussi un pendule de fréquence propre double de la première, que nous appellerons désormais « fondamentale ». Si nous sommes capables de donner un mouvement de fréquence triple, quadruple, etc., du fondamental, la corde entrera chaque fois en résonance, en formant trois, quatre, etc., fuseaux. Chaque mouvement propre s'appelle un « mode » ou un « partiel », sa fréquence étant multiple entier de la plus basse d'entre elles; on dit que les partiels sont harmoniques, et on les appelle souvent, par abréviation, les « harmoniques ».

Et maintenant un modèle théorique : au lieu d'une corde, imaginons une file de boules reliées les unes aux autres par des ressorts. Chaque boule ne peut se déplacer (peu importe par quel moyen, puisque nous ne sommes pas chargés de le réaliser) que perpendiculairement à la file. Ecartons la première boule de sa position d'équilibre; par le ressort, elle entraîne la deuxième, qui entraîne la troisième, et ainsi de suite. C'est la propagation de l'onde solitaire. La dernière boule est fixe; l'onde s'y réfléchit. Ces mouvements sont analysés de façon complète en appliquant à toutes les boules les lois de la dynamique : force égale masse multipliée par accélération. Imposons à la première boule un mouvement sinusoïdal et cherchons comment il faut choisir la fréquence pour que certaines boules de la file prennent des mouvements de grande amplitude. On trouve un grand nombre de solutions donnant à la file de boules les formes que nous avons décrites pour la corde; mais il y a une limite supérieure à la fréquence de ces parties.

Il est facile de voir pourquoi : ces mouvements présentent la double périodicité dans le temps et dans l'espace qui a été décrite à propos des ondes en général et de la lumière en particulier. Au cours des temps, toutes les boules marchent ensemble, passent ensemble par leur position d'équilibre (on appelle cela une « onde stationnaire »). Dans l'espace, chaque boule présente un certain retard de phase par rapport à la précédente; ce retard est le plus petit possible dans le mode fondamental, où la file entière ne représente qu'une demi-longueur d'onde; il est le plus grand possible quand chaque boule vibre en opposition avec celles qui l'encadrent. Impossible de faire plus; impossible de propager sur toute la file un mouvement de fréquence plus grande. On peut, en effet, toujours imposer à la première boule un mouvement aussi rapide que l'on veut; mais si l'on dépasse une certaine limite, l'autre boule refuse de fuir. Dans le langage des télécommunications, on dit que l'ensemble est un « filtre passe-bas ».

Evidemment irréalisable sous cette forme, le modèle de la file de boules n'est-il qu'un jeu de l'esprit? Non, si l'on sait s'en servir pour approcher par des perfectionnements successifs certains phénomènes réels. Si le nombre de boules est très élevé, on a une image plausible de la corde. Si leur nombre est limité, et si en même temps l'on a perfectionné les liaisons entre les boules pour tenir compte des résistances à la déformation dans le sens de l'écartement des boules entre elles et de la flexion de la file, on a fabriqué une molécule en chaîne dont la chimie organique est si riche. Si l'on place parallèlement les unes aux autres un nombre illimité (ce qui est toujours mathématiquement plus simple qu'avec un nombre fini) de chaînes liées maillon à maillon, on obtient un modèle de réseau cristallin, extrêmement utile pour décrire la dynamique si complexe des réseaux naturels. Prêtons un peu d'attention aux modèles moléculaires, par exemple à l'un des plus simples : celui où trois boules, une plus légère entre deux plus lourdes, constituent la molécule du gaz carbonique CO_2. Trois vibrations sont possibles : une transversale (du type détaillé ci-dessus), deux longitudinales, respectivement symétrique (les atomes O bougent en sens inverse) et antisymétrique (les atomes O bougent dans le même sens). L'expérience confirme pleinement tout cela et assigne aux « constantes élastiques » des valeurs cadrant bien avec tout ce qu'on sait par ailleurs des atomes en cause. Alors c'est un succès? Certes, mais partiel et provisoire.

Les fréquences propres que l'on observe dans tous les spectres ne sont pas seulement la fréquence fondamentale qu'on peut assigner aux divers modes de vibrations et à leurs harmoniques, mais se présentent comme des sommes et des différences de quantités que, faute d'en interpréter la nature, on a appelées des « termes », comme si la fréquence était une grandeur se débitant en tranches. Quant à l'absorption de la lumière, l'image de la résonance de pendule qui évoque une mise en branle progressive est certainement inadéquate, parce que le transfert de l'énergie à un atome se fait par quantités finies indivisibles.

Il s'agit d'une nouvelle mécanique, la « mécanique ondulatoire »; d'une nouvelle physique, la « physique des quanta », qui seront absolument nécessaires à l'interprétation des structures et des propriétés moléculaires et atomiques. Toutes les idées, tous les modèles que nous venons de voir ne sont que des étapes vers cette connaissance, étapes qu'il est indispensable de dépasser, mais qu'il est bien nécessaire de franchir.

Phot. Jidébé.

LA SCIENCE DES FORMES ET DES COULEURS

En terminant ce chapitre par l'étude des radiations visibles, juste hommage rendu à cet instrument irremplaçable de connaissance qu'est l'œil, nous allons retrouver notre vieil ami, le vecteur lumière, fort bien représentable par le vecteur de Fresnel.

Sa nature ondulatoire, bien établie, va nous permettre de concevoir aisément toute une catégorie de phénomènes dont l'étude constitue l'**optique physique.** Et cela en faisant abstraction, du moins pour le moment, de cette notion qui est venue pourtant papillonner déjà lorsque nous explorions l'ensemble du spectre électromagnétique : le concept des grains de lumière, des photons.

Plus tard, les photons apparaîtront tellement nécessaires qu'en retarder l'étude deviendrait un non-sens. Et ce sera l'**optique quantique.**

Mais conservons, pour un moment encore, une prudente réserve, et ne dédaignons pas le familier pinceau de lumière qui sait si habilement filtrer à travers le volet. Ce serait noire ingratitude que de refuser de chevaucher notre connaissance première : le rayon lumineux. Nous le suivrons donc à travers les instruments d'optique, nous le verrons se casser, se disperser et rendre l'âme dans un absorbant efficace. Et ce sera l'**optique géométrique.**

Il ne faut pas oublier que le fait « lumière » est toujours le même, quel que soit l'aspect sous lequel il se présente, que ce soit en optique physique, en optique quantique ou en optique géométrique.

Suivre pas à pas les conventionnelles divisions de l'optique géométrique et de l'optique physique amènerait si rapidement une telle dose d'ennui qu'il vaut mieux sans doute entrouvrir quelques portes, se donner une idée de quelques-uns seulement des aspects si divers du monde de la lumière.

L'électromagnétisme est un tout. Certaines méthodes, telle l'analyse harmonique, s'appliquent très largement à l'optique. Il n'en demeure pas moins que l'optique possède aussi une façon de voir les choses qui lui est propre, une manière spécifique de s'adapter. C'est ce que nous allons examiner.

Tout a commencé par la lumière...

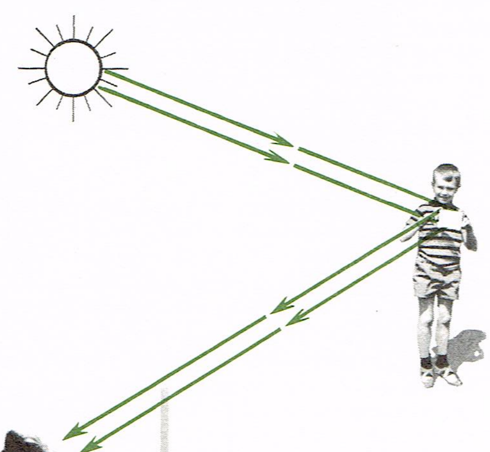
Par les belles soirées d'été, les chemins qui sillonnent les forêts du Jura sont envahis très tôt par de petites lumières vertes. Ce sont les vers luisants, les lampyres — pour utiliser l'appellation du naturaliste. Sa parure de nuit, dame ver luisant la doit à ses plaques abdominales, dont une substance complexe, la luciférine, s'oxyde à l'air et devient luminescente. Encore l'espèce de nos régions est-elle moins futée que ces lucioles américaines chez qui mâle et femelle communiquent, dit-on, au moyen de signaux particuliers : lui émet toutes les six secondes; elle répond au message au bout de deux secondes. Le temps de la réflexion, peut-être...

Ces lumières de la nature, on aurait tort de les considérer comme insignifiantes. Pendant la guerre du Pacifique, les Japonais savaient extraire de certains crustacés une poudre qui, mise dans l'eau, émettait une lueur bleuâtre plus intense que celle d'une allumette enflammée (un tel rayonnement a été nommé « lumière froide », toute l'énergie mise en jeu apparaissant sous forme de lumière).

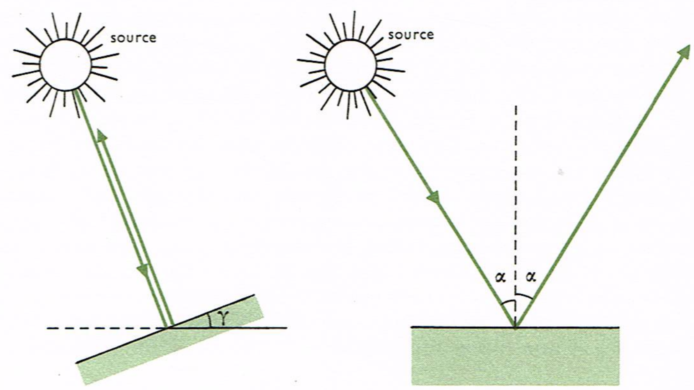
Voilà donc une source naturelle : la luminescence. Du fond de l'univers, les amas d'étoiles nous envoient, eux aussi, le rayonnement que leurs molécules et leurs atomes, bousculés, ont laissé échapper il y a de cela des milliers ou des milliards d'années. Pour étudier ces rayonnements, pour tenter de comprendre le merveilleux arc-en-ciel, pour mettre à profit l'extrême sensibilité du messager lumière, il a fallu se forger des outils, des instruments, des techniques.

réflexion et réfraction

Des expériences telles que celles qui sont résumées sur la figure ci-dessous on a déduit les lois de réflexion sur un miroir poli. Mais la surface du miroir peut être irrégulière, bien que, encore, en gros, réfléchissante. C'est le cas d'un mur blanc, d'une feuille de papier. La réflexion s'effectue rayon par rayon, si l'on peut dire, sur des miroirs élémentaires, orientés n'im-







Le Soleil, la perpendiculaire au miroir (normale) et la tête du personnage debout sont dans un même plan. Un faisceau de lumière — on dira un pinceau si la section est faible — tombant sur le miroir dont est armé le jeune garçon peut être braqué dans n'importe quelle direction suivant l'habileté de l'opérateur. On vérifie aisément que lorsque le miroir tourne d'un certain angle, le faisceau réfléchi tourne d'un angle double, du moins si la direction du faisceau incident est fixe. Si donc, au départ, le miroir renvoie la lumière d'où elle vient, faire tourner ce miroir d'un certain angle revient à faire tourner le rayon réfléchi d'un angle double. On établit ainsi une propriété générale de la réflexion : l'angle de réflexion est égal à l'angle d'incidence.

Phot. Jidébé.

porte comment. Les rayons réfléchis prennent alors toutes les directions. On dit qu'il y a réflexion diffuse.

Revenons à notre miroir de « bonne qualité », Regardée dans la glace la main droite apparaît comme une main gauche. On dit que l'image est énantiomorphe! En termes moins choisis : la réflexion équivaut à une symétrie par rapport au plan du miroir. Ainsi, ne peut-on pas superposer l'objet et l'image.

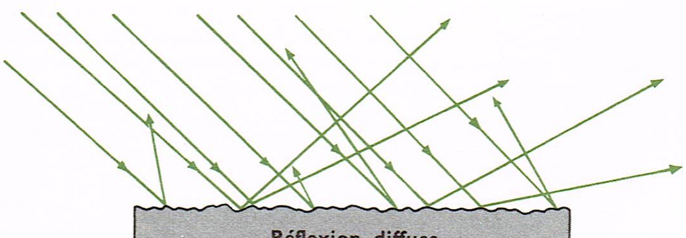
Maintenant, quelle est la nature de cette image? Peut-on la recueillir sur un écran? Non, on sait bien que derrière le miroir il n'y a rien. Et pourtant chacun des points de l'objet semble à la même distance du miroir que le point image correspondant. On peut photographier le pont qui se mire dans l'eau calme de la Loue, on n'ira pas chercher un autre pont au fond de la rivière; on dit que l'image est virtuelle.

La surface de l'eau est un miroir plan. Les miroirs grossissants sont sphériques; leur concavité est tournée vers le sujet. On pourrait construire l'image fournie par un système grossissant en considérant chaque point objet et en utilisant les lois de Descartes, et l'on constate qu'elle est virtuelle.

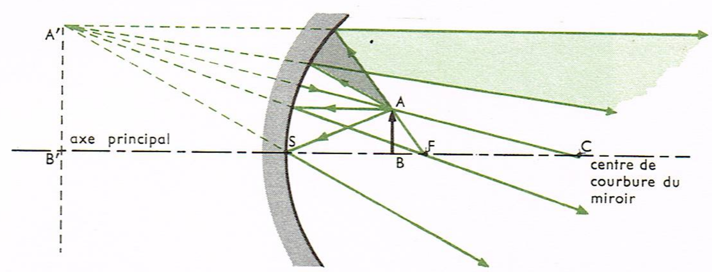
Pour déterminer la position et la grandeur de l'image d'un objet donné, on peut, plus simplement, se fonder sur quelques remarques simples. Par exemple celle-ci : les rayons issus d'une source très éloignée (on parlera de source à l'infini : c'est le cas du soleil) convergent pratiquement en un point. Ce point, le foyer, est situé à égale distance du sommet du miroir et de son centre de courbure (centre de la sphère dans laquelle on aurait découpé le miroir). La réciproque est vraie : une source placée au foyer permet de réaliser un faisceau parallèle; la propriété est utilisée dans les projecteurs et dans les phares des automobiles. Nous disposons donc de trois catégories de rayons pour construire l'image de l'objet AB.



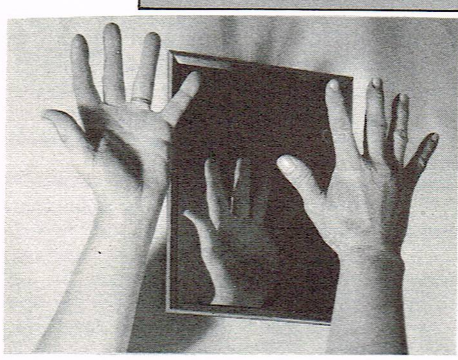
Miroir plan dans la nature.



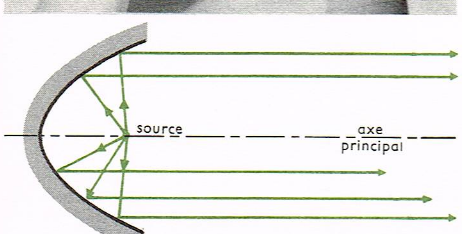
Réflexion diffuse.



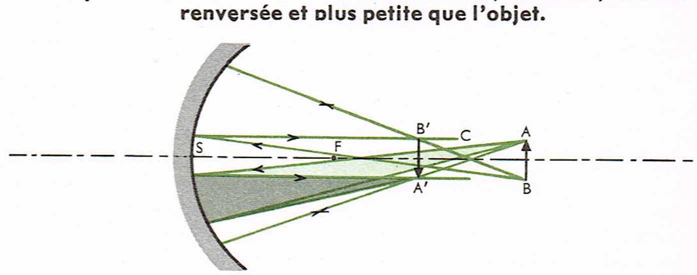
Miroir sphérique concave.
Les trois rayons issus de A semblent finalement provenir d'un point A' situé en arrière du miroir : A' est un point image virtuel. A chacun des points de l'objet correspond un point conjugué, son image.
Dans le cas de la figure ci-dessus, l'image est virtuelle et droite, c'est-à-dire tournée dans le même sens que l'objet. Elle est d'autant plus agrandie que l'objet est plus près du foyer, côté miroir. Il n'en est pas toujours ainsi : lorsqu'on recule l'objet au-delà du foyer (nous nous plaçons, par exemple, dans la situation de la figure ci-dessous), l'objet AB s'éloigne, l'image devient réelle située en avant du miroir, elle peut être recueillie sur un écran. Mais, cette fois, elle est renversée et plus petite que l'objet.



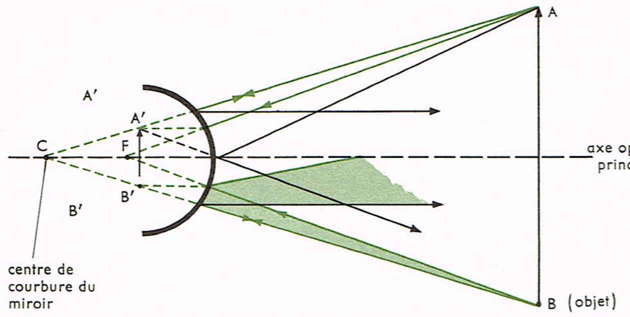
La main droite apparaît comme une main gauche dans le miroir.



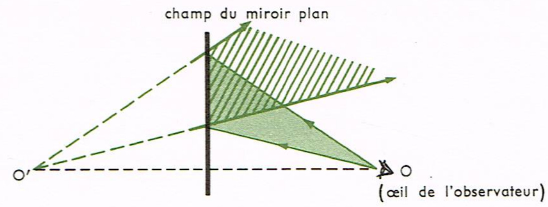
Miroir parabolique, dont les phares des automobiles constituent l'une des applications courantes.

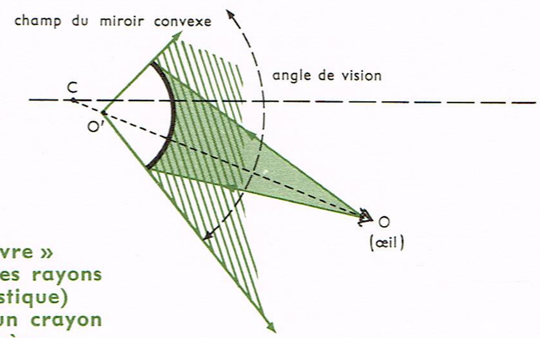


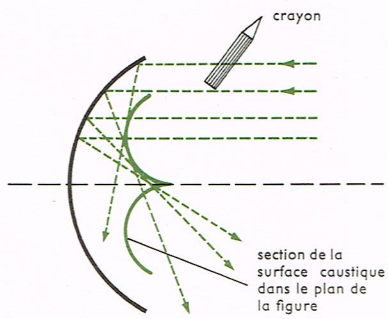
Phot. Larousse.



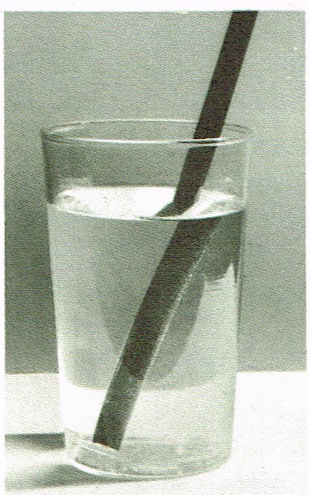
Construction de l'image donnée par un miroir sphérique convexe.







On peut « suivre » l'enveloppe des rayons réfléchis (caustique) au moyen d'un crayon que l'on promène dans tout le champ; son ombre vient successivement en tangence de tous les points de la caustique.



Réfraction : le bâton plongé dans l'eau paraît brisé.

quelques applications

Les rétroviseurs montés sur certaines voitures, les grands miroirs placés quelquefois au coin des rues permettent de voir un champ très étendu. Il s'agit de miroirs sphériques convexes. D'un objet éloigné, ils donnent une image virtuelle droite plus petite que l'objet. Il n'est pas étonnant que le champ angulaire d'un tel miroir soit beaucoup plus étendu que celui d'un miroir plan de mêmes dimensions. On conçoit donc leur utilité.

Ces miroirs qui paraissent si simples sont-ils des instruments parfaits? L'étude des imperfections (des aberrations) des systèmes optiques nécessiterait de longs développements. Néanmoins, il est facile de se rendre compte de l'effet de l'une d'entre elles : l'aberration sphérique; lorsque l'ouverture d'un miroir est importante, c'est-à-dire lorsque son diamètre est grand par rapport à son rayon de courbure (miroir très bombé) — disons une fois et demie, par exemple — les propriétés de convergence d'un faisceau ne restent plus tout à fait celles sur lesquelles nous comptions tout à l'heure (1); ainsi, tous les rayons d'un faisceau parallèle à l'axe ne viennent-ils pas, après réflexion, converger au foyer. Ils présentent pourtant encore une qualité commune : ils restent tangents à une même surface, la « caustique ». Ce phénomène est observable au fond d'une tasse ou, mieux, d'un bol vide, lorsqu'on le place sous la lampe avec une incidence oblique. Il est possible de réduire cette aberration en diminuant l'ouverture du système, c'est-à-dire en ne conservant que les rayons cheminant au voisinage de l'axe optique (c'est ce qu'on appelle les conditions de l'approximation de Gauss). Mais la faible quantité de lumière utilisée réduit alors sensiblement les possibilités du système. Le plus souvent, il est rentable de réaliser des corrections de courbure (abandon de la forme sphérique au profit du paraboloïde), d'interposer des lames correctrices (lames de Schmidt), ou, enfin, de se résoudre à compliquer le système par l'association de plusieurs éléments simples : miroirs, lames, lentilles. C'est la solution adoptée dans la plupart des instruments d'optique.

la réfraction

Il est facile de présenter des photographies illustrant un tel phénomène. Les rayons situés dans le second milieu ont changé de direction à la surface de séparation : ils se sont réfractés. En faisant varier l'angle d'incidence, on constate que le rapport des sinus des angles d'incidence et de réflexion est constant (loi de Descartes-Snell); *n* s'appelle « indice de réfraction » et caractérise le second milieu par rapport au premier.

---

(1) Supposons un faisceau divergent issu d'un point objet : lorsque après réflexion, ou, de façon générale, après passage par un instrument d'optique, le faisceau ne converge pas en un point, on dit que le système présente un défaut de stigmatisme. Il y a stigmatisme rigoureux lorsque l'image d'un point est un point (au sens strictement géométrique).

Phot. Larousse.

La réfraction correspond à un changement de vitesse de la lumière lors du passage d'un milieu à un autre (1), ce qu'indique la figure où l'on a tracé les rayons et les ondes des faisceaux incident et réfracté. Si l'on se souvient de ce qui a été dit des ondes et du déplacement des trains d'ondes, le transport de l'énergie lumineuse suivant la direction du rayon devient une notion très intuitive; les ondes sont perpendiculaires aux rayons, et discuter à partir de l'un ou de l'autre de ces deux concepts reste ici équivalent. Donc, au passage dans le second milieu, supposé, comme dans la figure, d'indice n_2 supérieur à n_1, il y a freinage de la lumière : v_2 est inférieur à v_1. Si l'on suppose que les deux milieux présentent tous deux les mêmes propriétés dans toutes les directions (homogénéité et isotropie), il faut bien que l'orientation des ondes réfractées soit différente de celle des ondes incidentes, puisque au bout de chaque période T la distance parcourue par une crête n'est plus v_1T, mais v_2T (inférieure à v_1T); ainsi doit-il y avoir resserrement des creux et des crêtes, ce qui n'est possible qu'au prix d'un changement d'orientation. Pour une valeur donnée de l'indice, il existe une valeur limite de l'angle de réfraction r, pour laquelle l'angle i vaut 90° (incidence rasante), puisque sin i doit être inférieur à 1, donc sin r inférieur à $1/n$, et $r <$ arc sin $(1/n)$;
soit $r < 48°35$ pour la réfraction eau/air ($n \sim 1{,}33$),
$r < 41°48$ pour la réfraction verre/air ($n \sim 1{,}5$).

Réciproquement, lorsque la lumière venant du second milieu plus réfringent arrive sur le premier sous un angle supérieur à cette valeur limite, elle ne pourra pas émerger : on aura affaire à une réflexion totale. C'est ce qu'illustrent les fontaines lumineuses à l'intérieur desquelles le pinceau lumineux subit des réflexions totales sans jamais pouvoir en sortir.

les irisations

Une autre conséquence de la réfraction est la dispersion des couleurs par le prisme. Au moyen d'un simple prisme de verre, Newton a montré en 1666 qu'un faisceau de lumière blanche s'étalait en une gamme de couleurs allant du rouge au violet (spectre).

Que se passe-t-il? La lumière blanche est la superposition, la synthèse, de plusieurs couleurs en proportions déterminées.

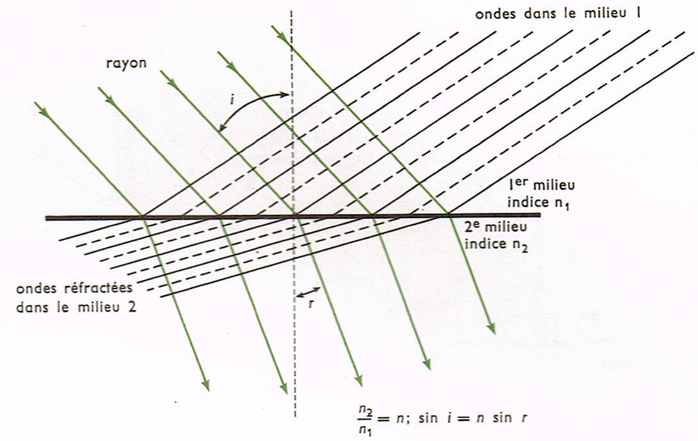
A chaque couleur correspond une longueur d'onde, donc une vitesse de propagation v (la relation $\lambda = vT$ a déjà été vue précédemment). Ainsi, le prisme, milieu réfringent, ne se comporte pas de la même manière avec les différentes couleurs; dans le vide, au contraire, toutes les ondes électromagnétiques se propagent à la même vitesse. Aux plus grandes longueurs d'onde (rouges) correspondent les plus grandes vitesses; le « rouge » subira donc une réfraction, c'est-à-dire finalement une déviation, moindre que celle du violet : son indice de réfraction sera moindre.

l'arc-en-ciel

Occupons-nous maintenant de ce curieux phénomène : l'arc-en-ciel. Noé le vit briller au moment où le dernier rideau de pluie fuyait devant le soleil revenu. Newton l'expliqua au moyen de son prisme, Descartes aussi en fit la théorie.

La lumière du Soleil arrivant sur chacune des gouttes de pluie tombant du nuage subit une première réfraction, puis une réflexion interne, d'autant plus importante que l'angle d'incidence est grand (à la limite on retrouve évidemment le

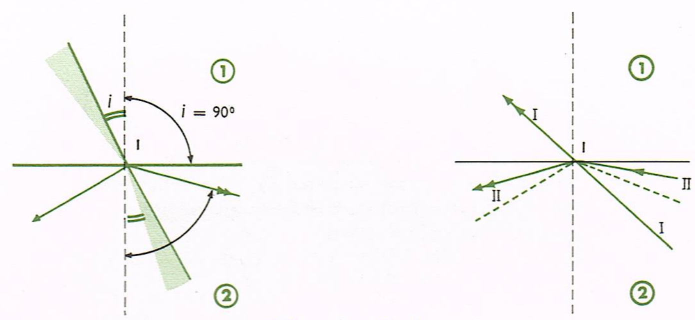
(1) L'indice de réfraction absolu d'une substance transparente est égal au rapport c/v de la vitesse de la lumière dans le vide à celle dans le milieu considéré. L'indice de réfraction relatif d'une substance par rapport à une autre, n, est égal au rapport des indices absolus, c'est-à-dire, en fait, des vitesses de la lumière dans les milieux considérés : $n = \frac{n_2}{n_1} = \frac{cv_1}{cv_2} = \frac{v_1}{v_2}$. L'indice de l'air, $n = 1{,}000\,293$, diffère fort peu de celui du vide; donc l'indice par rapport à l'air est pratiquement égal à l'indice du vide absolu.



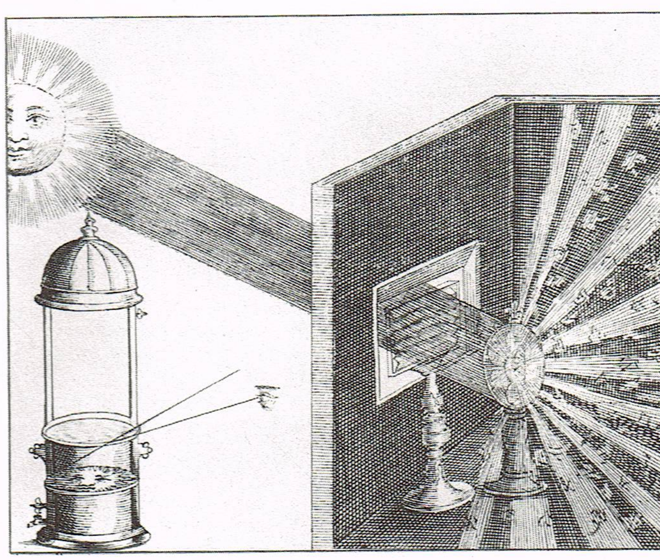
Passage d'un faisceau lumineux d'un milieu à un autre.

- - - - - - - - *minimum d'amplitude de l'onde (creux)*

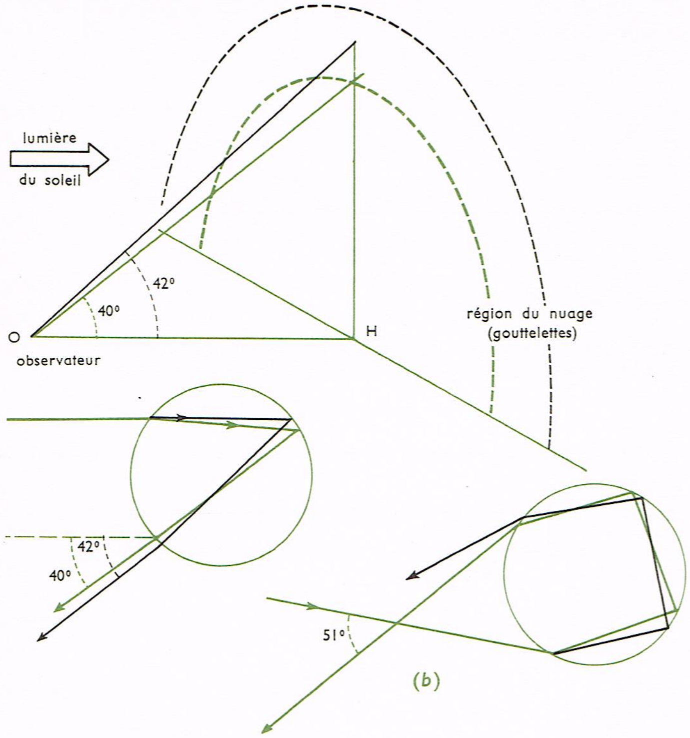
——————— *maximum d'amplitude de l'onde (crête)*



Réfraction limite.
Les rayons arrivant sur la surface de séparation des deux milieux ne peuvent pas toujours traverser cette surface.



Cette gravure montre combien les idées sur la réfraction de la lumière étaient encore vagues au XVIIe siècle. L'expérience de Newton montra pourtant qu'en traversant un prisme la lumière solaire s'étalait en un spectre de lumières monochromatiques allant du rouge au violet.



Principe de l'arc-en-ciel.
Le rayon lumineux subit généralement une seule réflexion à l'intérieur de la goutte de pluie (*a*). S'il y est réfléchi deux fois (*b*), il se produit un arc-en-ciel double.

phénomène de réflexion totale), enfin une seconde réfraction. On peut montrer, compte tenu de l'indice de réfraction de la goutte, que le faisceau de lumière émergent atteint son intensité maximale lorsque son inclinaison moyenne est de 42° environ par rapport à l'incident. D'autre part, à l'instar du prisme de tout à l'heure, la goutte se comporte comme un système dispersif : les radiations violettes, de plus courte longueur d'onde, sont plus déviées que les rouges. Le rayon violet ou bleu est donc moins incliné sur l'horizontale OH que le rayon rouge. Ainsi, l'observateur qui contemple les nuages en train de se résoudre verra le rouge au-dessus du bleu. (Nous nous sommes placés par raison de commodité dans un plan vertical, mais le raisonnement reste bien entendu valable pour tous les plans ayant en commun la ligne OH.) Aux cônes de rayons rouges, plus ouverts que ceux de rayons bleus, correspondent les régions extérieures — rouges — de l'arc-en-ciel.

Quelquefois même apparaît un deuxième arc-en-ciel, plus pâle que le premier. Deux réflexions successives ont alors eu lieu à l'intérieur de chacune des gouttes. L'intensité maximale correspond alors à des rayons émergents faisant un angle de 51° environ avec la lumière incidente. Dans un arc secondaire, extérieur à l'arc principal, les couleurs sont interverties, et les intensités moindres, du fait de la double réflexion.

le mirage

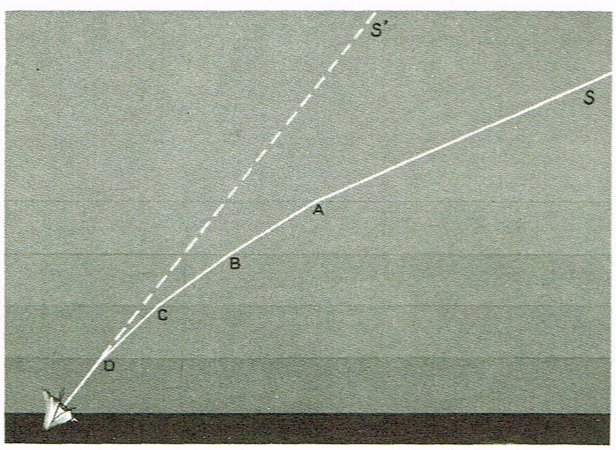
Lorsqu'on regarde par-dessus un radiateur chaud, il semble que les objets ondulent. Par une journée brûlante d'été, quand le soleil a fortement réchauffé la route, quel automobiliste n'a pas eu l'impression d'un miroir étalé devant lui?

C'est qu'au voisinage d'une surface chaude, l'air, moins dense, est moins réfringent (l'indice *n* est plus faible) que dans les couches plus élevées. Cette inégalité se traduit par une courbure des rayons lumineux au voisinage de la surface.

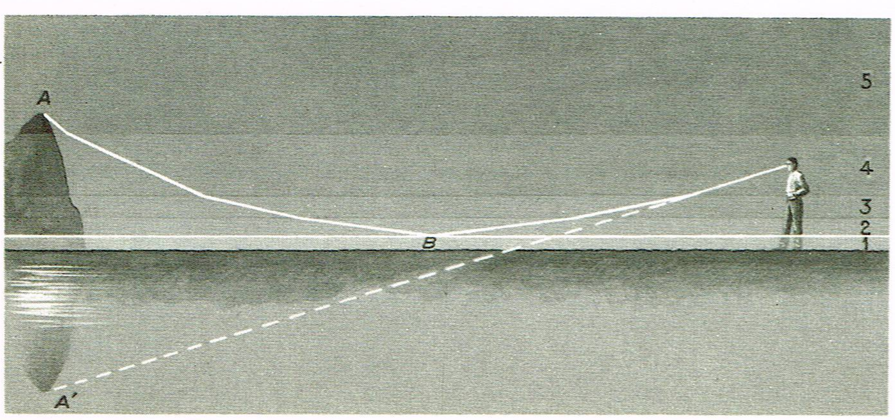
L'œil ne sait tenir compte de cette courbure; il s'imagine que l'objet est dans la direction de la tangente au rayon.

Sur la route, on ne verra sans doute pas tout à fait, de la deuxième voiture, la première retournée, mais à grande distance et en l'absence de tourbillons locaux le phénomène est plus net, et pour le voyageur assoiffé du désert l'illusion d'une nappe liquide salvatrice devient complète (la chaleur et l'imagination aidant).

Un phénomène inverse se produit lorsque des couches d'air froid traînent au voisinage du sol. Les régions basses, plus denses donc plus réfringentes, rabattent les rayons lumineux. Ainsi, lorsque vers Maisons-Alfort le voyageur du « Mistral » découvre la tour Eiffel, celle-ci apparaîtra-t-elle plus haute qu'à l'accoutumée si une vague d'air froid déferle sur Paris. Cela n'est pas très important; ce phénomène aura même vraisemblablement une amplitude assez faible, et le voyageur ne s'en apercevra pas. Beaucoup plus délicate devient la correction que doit faire un astronome lorsqu'il détermine la hauteur d'une étoile; une erreur d'angle de 1° n'est plus alors admissible, ce qui oblige l'observateur à faire des corrections, calculées une fois pour toutes en fonction des conditions atmosphériques.



Principe de la réfraction atmosphérique.
Le rayon venant de S, au terme de multiples réfractions (A, B, C, D), parvient à l'œil avec une inclinaison telle qu'il semble provenir de S'.



Explication d'un mirage.
Les couches d'air (1 à 5) sont ici d'autant plus chaudes qu'elles sont proches du sol.

Le rayon issu de A, progressivement réfracté, change de sens en B et parvient à l'œil comme s'il avait son origine en A'. On voit donc au sol une image inversée de la montagne. Si l'air était plus froid à proximité du sol qu'en altitude, la courbe serait inversée, et le point B se trouvant alors dans la couche 5, l'œil verrait la montagne plus haute qu'elle ne l'est.

lentilles, loupes, objectifs

Un morceau de substance réfringente placé dans un milieu d'indice plus faible dévie les rayons lumineux. Lors de leur fabrication, les lentilles sont taillées suivant des formes géométriques élaborées de telle manière que les faisceaux de lumière convergent, ou divergent, suivant les lois presque aussi simples que celles de la réflexion. Les lentilles sphériques présentent une symétrie de révolution autour de l'axe de la sphère dans laquelle elles ont été découpées.

Les lentilles cylindriques ne présentent pas, bien entendu, les mêmes propriétés suivant la direction de l'axe du cylindre dans lequel elles sont taillées et suivant la direction perpendiculaire. Cela est mis à profit dans la correction d'un défaut particulier de l'œil : l'astigmatisme. On détecte un tel défaut lorsque le patient, en regardant les rayons d'une roue, les voit apparaître nets suivant une direction et flous suivant une autre direction, souvent perpendiculaire. Des verres spéciaux (toriques) corrigent à la fois l'astigmatisme et les autres défauts que peut présenter un même œil.

L'instrument d'optique le plus courant est certainement la loupe. La plus élémentaire est constituée par une seule lentille, convergente, donnant une image virtuelle plus grande que l'objet et située du même côté (1). Il semble que l'image apparaîtra d'autant plus grande que l'on s'arrangera pour l'éloigner. En fait, comme pour les miroirs, il existe un foyer, et l'image de tout objet placé en ce point est rejetée à l'infini, agrandie indéfiniment. L'œil n'a plus besoin d'accommoder pour voir nettement cette image, mais, pour profiter au maximum de l'agrandissement, il lui faut la regarder sous un angle aussi grand que possible : l'œil doit alors prendre place au voisinage du foyer image F′ de la loupe. Pratiquement, on place la loupe à une distance de l'objet telle que celui-ci soit juste sur le point de disparaître (un éloignement plus grand en ferait une image réelle, qui se formerait du côté de l'observateur) et l'on rapproche l'œil de la loupe.

Les systèmes optiques (objectifs, oculaires, par exemple) sont des combinaisons de lentilles simples — convergentes ou divergentes (systèmes dioptriques) —, associées quelquefois à des miroirs (on les appelle alors « systèmes catadioptriques »). Ainsi, l'oculaire d'une lunette ou d'un télescope permet-il d'examiner l'image réelle fournie par l'objectif et située en avant de l'appareil. Cet oculaire joue simplement le rôle d'une loupe, traduisant en image virtuelle encore plus agrandie l'image intermédiaire. L'une des caractéristiques essentielles de la lunette, son grossissement, est le rapport des angles sous lesquels l'œil voit l'objet à travers l'instrument et à l'œil nu.

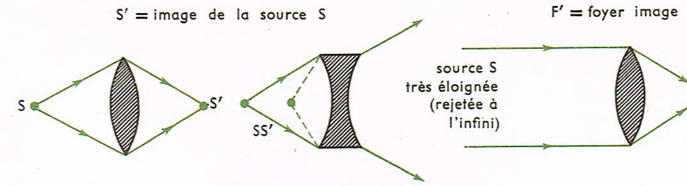
En outre, il est nécessaire qu'une lunette terrestre offre une image droite, et non pas renversée. Pour cela, elle est munie d'un système redresseur, dont le type classique est essentiellement un système de prismes. Deux lunettes de grossissement relativement faible (6, 8 ou 16, par exemple), convenablement juxtaposées et réunies, constituent une paire de jumelles.

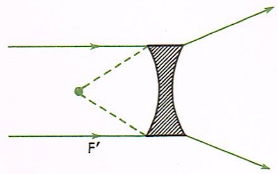
Les lunettes astronomiques ne diffèrent des lunettes terrestres que par l'absence de redresseur d'image et aussi par les dimensions.

---

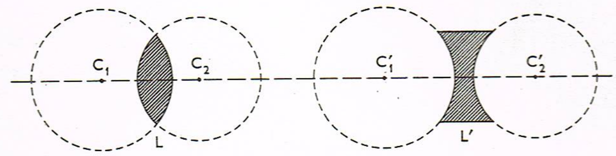
(1) On ne reprendra pas ici les lois, même élémentaires, de formation des images comme il avait été fait pour les miroirs. Une seule « formule », que connaissent bien les candidats au baccalauréat, peut être rappelée. Rappelons la formule, qui relie — avec une bonne approximation dans le cas où l'on a affaire à une lentille mince — la distance focale f (ou encore la quantité $1/f$, appelée convergence de la lentille) aux distances x de l'objet et x' de l'image comptées algébriquement à partir de la lentille : $1/x' = 1/x + 1/f$, le grandissement, c'est-à-dire le rapport des grandeurs de l'image et de l'objet a pour valeur x'/x. Il existe d'autres groupes de relations qui ne semblent pas indispensables à la compréhension de la suite du texte.

x et x' et le grandissement g sont reliés par les deux relations : $\frac{1}{x'} = \frac{1}{x} + \frac{1}{f}$ (voir distance focale) $g = \frac{x'}{x}$ du moins lorsque l'on a affaire à une lentille mince (et qu'on est dans les conditions des approximations de l'optique géométrique).



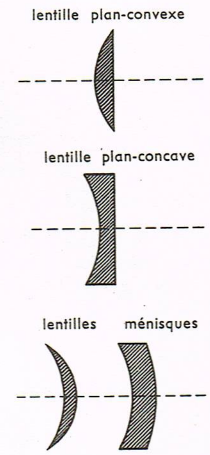


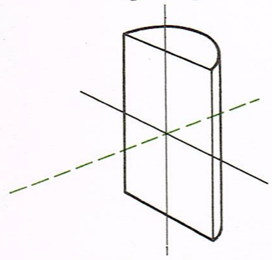
Comportement d'une lentille convergente, puis d'une lentille vis-à-vis d'un faisceau quelconque puis d'un faisceau de rayons parallèles.



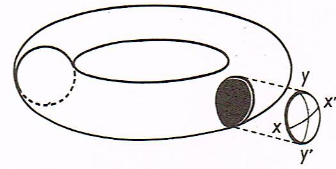
La lentille sphérique biconvexe L est représentée par la partie commune à deux sphères de centres C_1 et C_2.

La lentille sphérique biconcave L′ correspond à une région comprise entre deux sphères de centres C'_1 et C'_2.

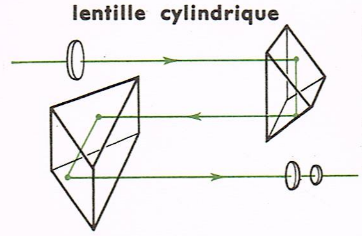




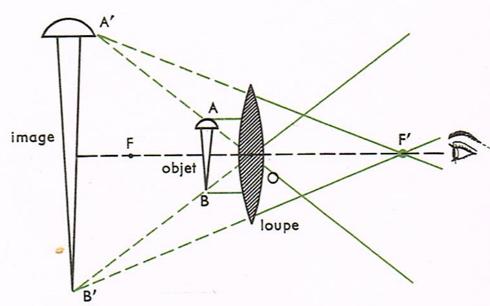
lentille cylindrique

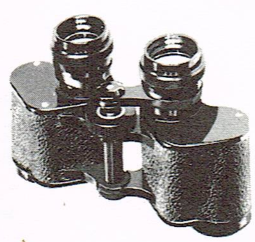


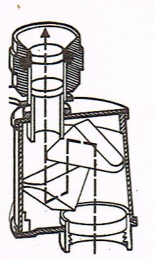
Découpage dans un tore d'une lentille pour astigmate. La combinaison des propriétés d'une lentille cylindrique et d'une lentille sphérique est réalisée par une lentille torique, c'est-à-dire découpée dans un volume semblable à un pneu dont les courbures suivant les directions principales x et x' et y et y' sont différentes.



lunette avec système redresseur à prisme







jumelles à prisme

Image donnée par une loupe; en bas l'objet est au foyer : les rayons AO et F′A′ sont parallèles.

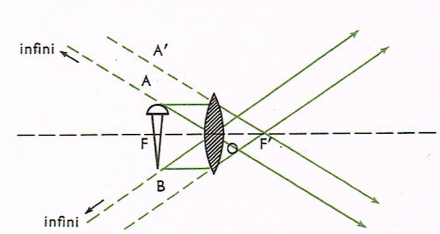
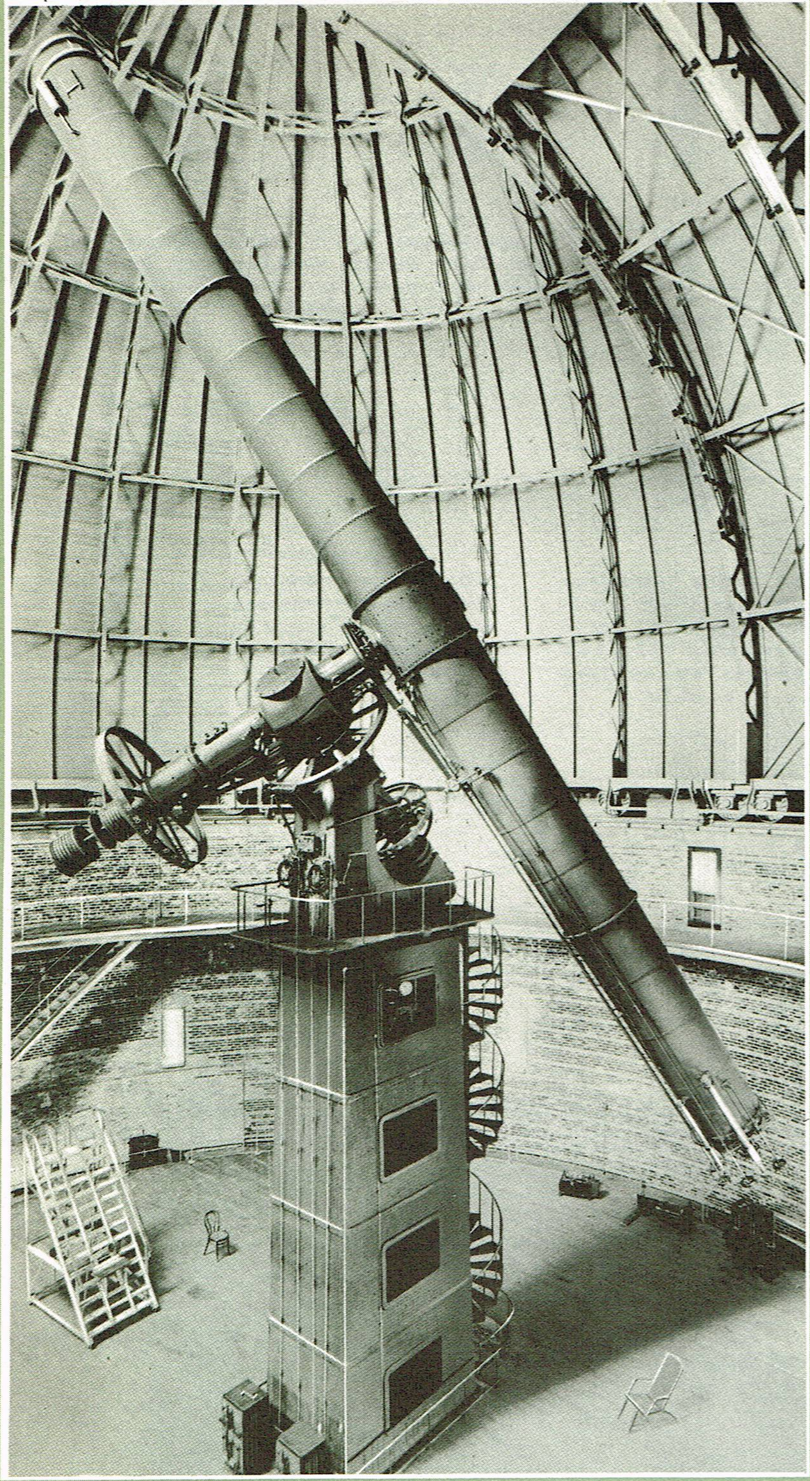
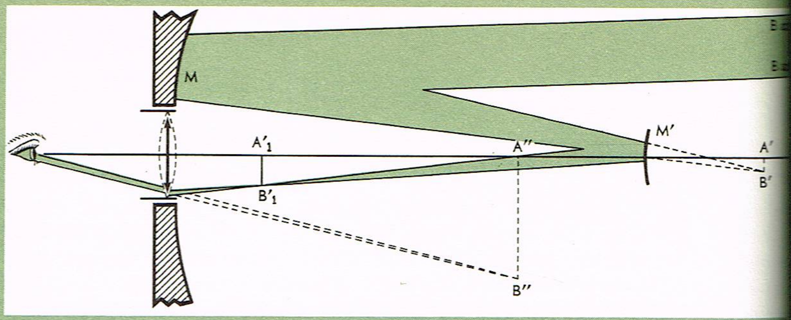


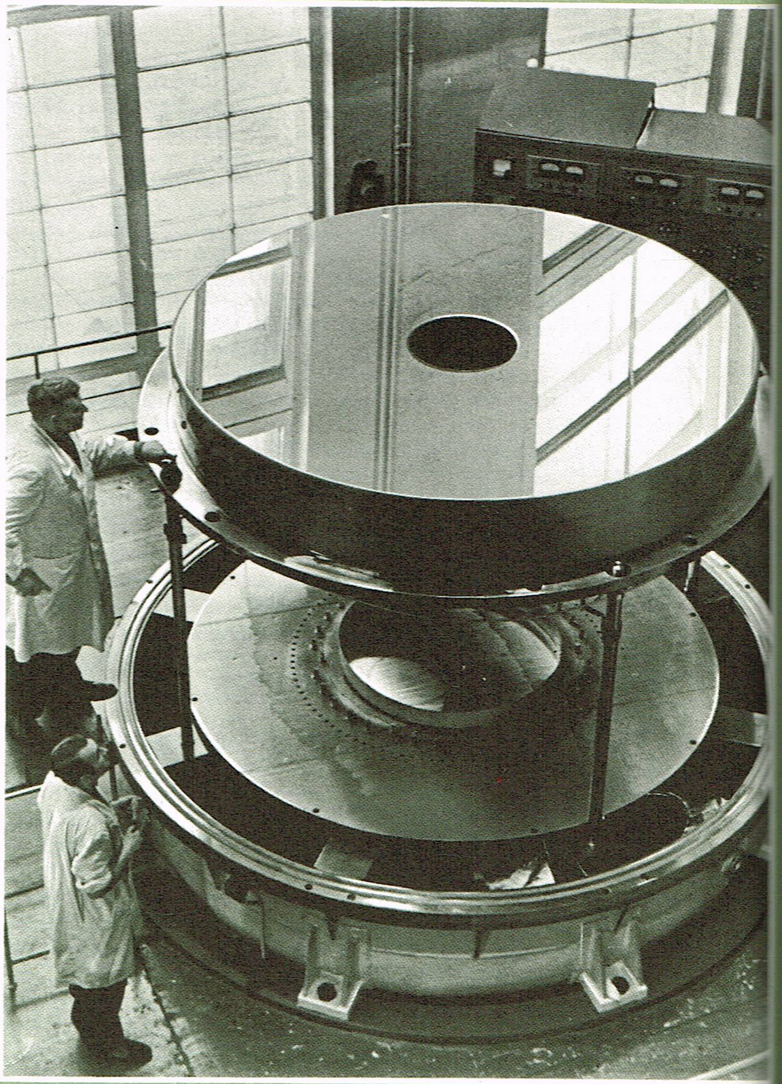
Schéma montrant le fonctionnement d'une lunette d'approche.





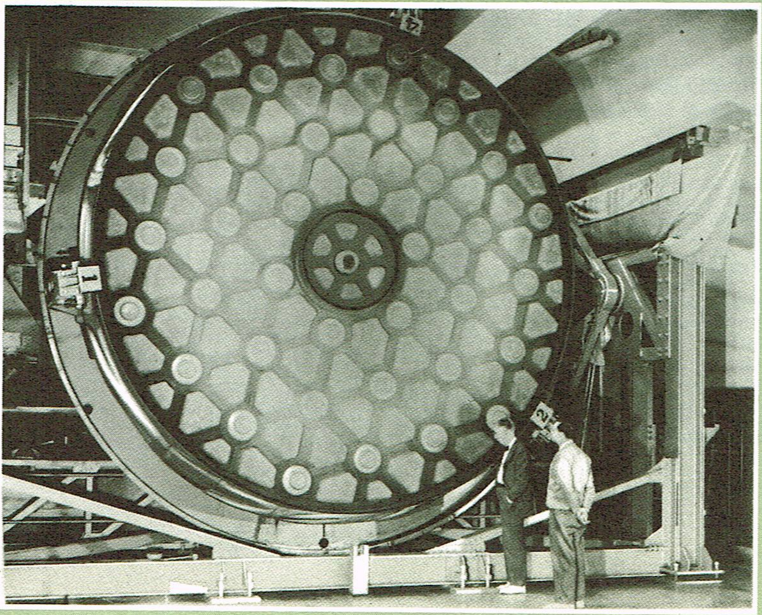


télescope à vision directe



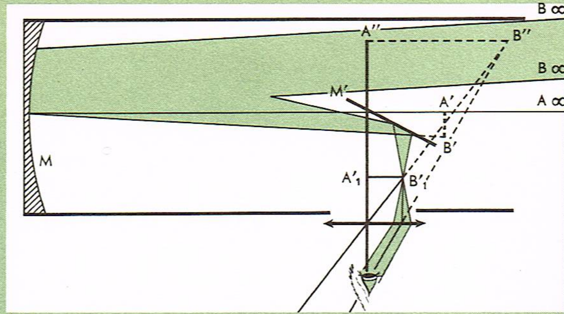
Aluminage sous vide d'un miroir de 2,60 m de diamètre, pesant 4 tonnes.

La grande lunette de l'observatoire Yerkes (à Williams Bay, dans l'État de Wisconsin). Caractéristiques de cet instrument, qui est le plus grand réfracteur jamais construit : ouverture, 102 cm; longueur focale, 19,3 m; entrée en service, 1897.

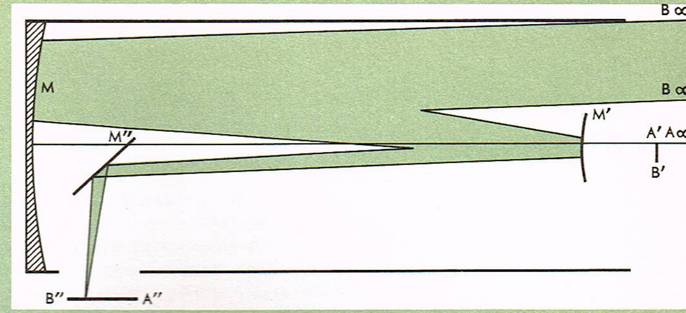


Le miroir du télescope du Mont-Palomar, le plus grand du monde, ne mesure pas moins de 5 m de diamètre.

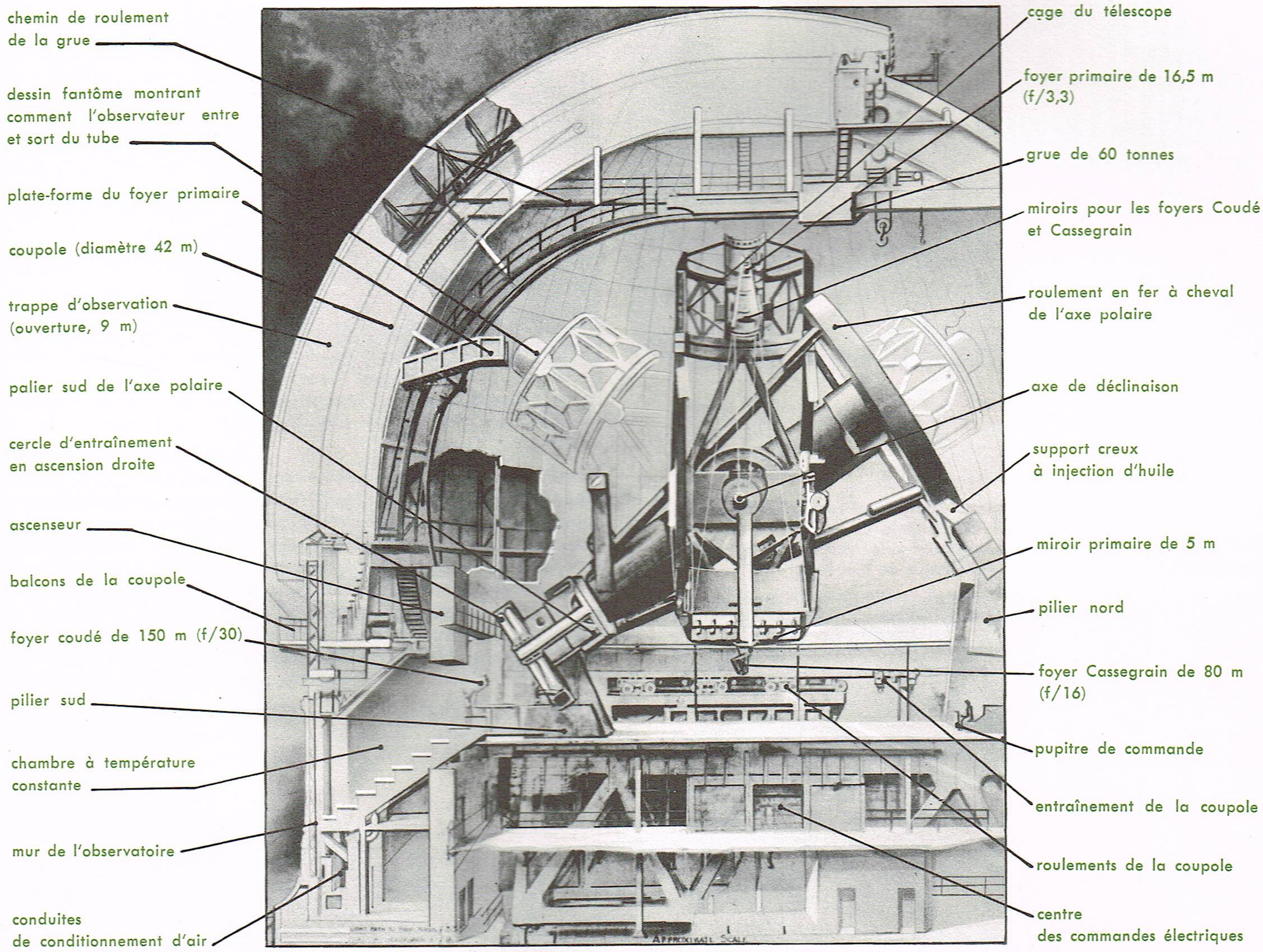
Marche des rayons dans les trois variantes usuelles du télescope :
B∞, faisceau parallèle;
M, miroir objectif;
M', miroir réfléchissant;
B', convergence des rayons du faisceau B;
B'₁, image substituée à B' par le miroir M;
l'image réelle A'₁ B'₁ est vue en A'' B'' au moyen de l'oculaire.



télescope à vision latérale



télescope du Mont-Wilson



Coupole du télescope géant de l'observatoire du Mont-Palomar.

Le diamètre de l'objectif d'une bonne petite lunette d'amateur fait 8 à 10 cm; celui de la lunette de l'observatoire de Meudon est de 83 cm (distance focale : 16,2 m; grossissement utilisé : de l'ordre de 1 600); celui de la lunette de Yerkes vaut 102 cm.

Un télescope, appareil mettant en œuvre à la fois miroirs et lentilles, est représenté ici dans ses différents montages. L'objectif est un miroir concave, pas tout à fait sphérique, mais parabolique, ce qui permet de réduire l'aberration de sphéricité (les rayons parallèles, écartés de l'axe, ne convergent pas au foyer comme il a été vu ci-dessus lorsqu'il a été question des caustiques). Le miroir est métallisé, recouvert d'une mince pellicule d'argent ou d'aluminium, obtenue par évaporation sous vide, qui permet un pouvoir réflecteur élevé. Pour se donner une idée de l'importance des miroirs de télescopes il faut se rappeler que celui de l'observatoire de Yerkes pèse 2 370 kg et mesure 2 m de diamètre; le tube qui le supporte (20 m de long) pèse à lui seul plus de 26 tonnes. Le miroir du télescope du Mont-Palomar a 5 m de diamètre, les défauts de sa surface ne dépassent pas 0,000 05 mm. Qu'on juge du travail titanesque exigé par la taille, qui a duré sept ans, y compris le temps de refroidissement de la masse (une technique plus rapide a été mise au point en France par Couder à l'occasion de la taille du télescope de Saint-Michel-l'Observatoire).

Mais ces instruments géants, qui recueillent la lumière sur plusieurs mètres carrés pour la concentrer sur la pupille de l'œil ou la plaque photographique, révèlent les plus lointains secrets de l'univers (voir tome I).

Doc. Institut de technologie de Californie.

Phot. Ellerman, Bureau soviétique d'information.

le microscope

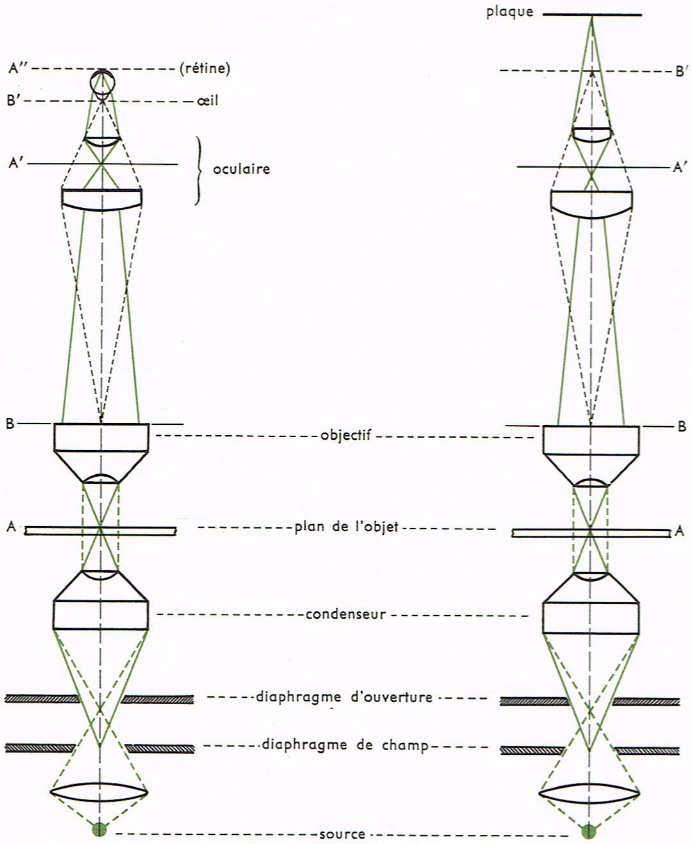
L'objet est vu sous un angle d'autant plus grand que la loupe est plus convergente. Or, la distance focale ne pouvant pas être réduite au-dessous d'une certaine valeur, la convergence (ou, ce qui revient au même, la puissance) ne saurait être accrue indéfiniment. L'examen des objets de plus en plus petits situés à faible distance (à l'inverse de ce qui se passe pour une planète observable à la lunette) ne sera plus possible au moyen d'un dispositif élémentaire.

On procédera en deux temps. Une première image, réelle, est donnée par l'objectif du microscope, combinaison de lentilles calculées de telle manière que cette image intermédiaire soit aussi bonne que possible, aussi semblable à l'objet que l'on puisse souhaiter (il s'agira alors d'un système corrigé pour différents types d'aberrations, de haute « définition », c'est-à-dire obtenant une image dans laquelle chaque détail est reproduit fidèlement). L'image intermédiaire se comporte comme un objet pour l'œil, qui l'examine à travers cette loupe complexe qui est l'oculaire. La coupe d'un microscope moderne rend compte de la complexité des différents éléments.

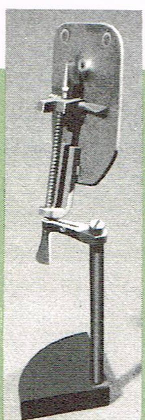
Les techniques de la microscopie optique sont capables de performances encore plus belles que celles qu'illustrent les photographies ci-dessous. On peut rendre visibles des contrastes presque imperceptibles, des objets dont les dimensions sont inférieures au cent-millième de millimètre. Mais il faut faire appel aux « microscopes à contraste de phase » (v. tome II), dont le fonctionnement, inexplicable par les lois de l'optique géométrique, fait appel aux propriétés ondulatoires de la lumière.

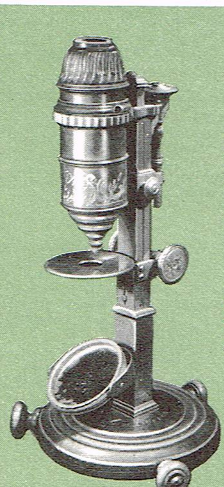
Comment chiffrer les qualités de grossissement d'un microscope? On a vu que l'objectif donne un « agrandissement » sous forme d'une image réelle dont les dimensions peuvent être rapportées à celles de l'objet. A son tour, l'oculaire se comporte comme une loupe et grossit pour l'œil l'image intermédiaire (le grossissement d'un microscope s'exprime par le produit du grandissement de l'objectif par le grossissement de l'oculaire). Peut-on alors, en calculant et en utilisant des systémes de plus en plus complexes et de plus en plus précis, augmenter toujours davantage le grossissement de l'appareil?

Certes non. Pour l'expliquer, il faut retourner à l'aspect ondulatoire de la lumière. Nous étions prévenus : l'optique des rayons, géométrie trop simpliste, n'est qu'une première approximation.



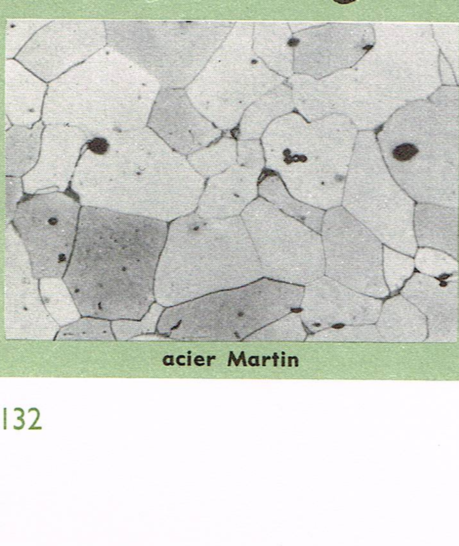
Principe du microscope en observation visuelle : A A' A'' sont des points conjugués, B B' sont également conjugués en microphotographie (éclairage de Köhler).



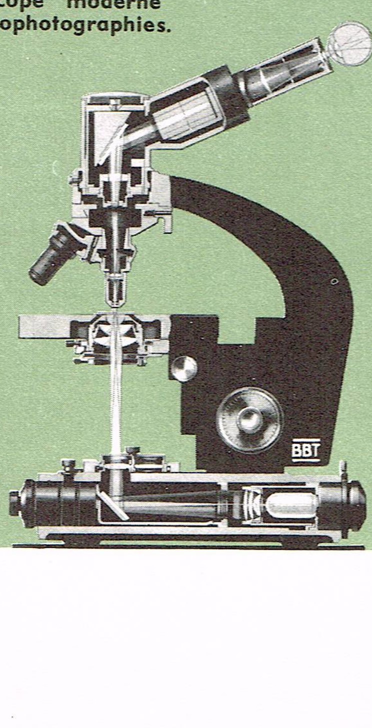


Les microscopes de Van Leeuwenhoek *(à gauche)* et de Buffon *(ci-contre)*.

***Ci-dessous,* détail d'un microscope moderne et microphotographies.**



acier Martin

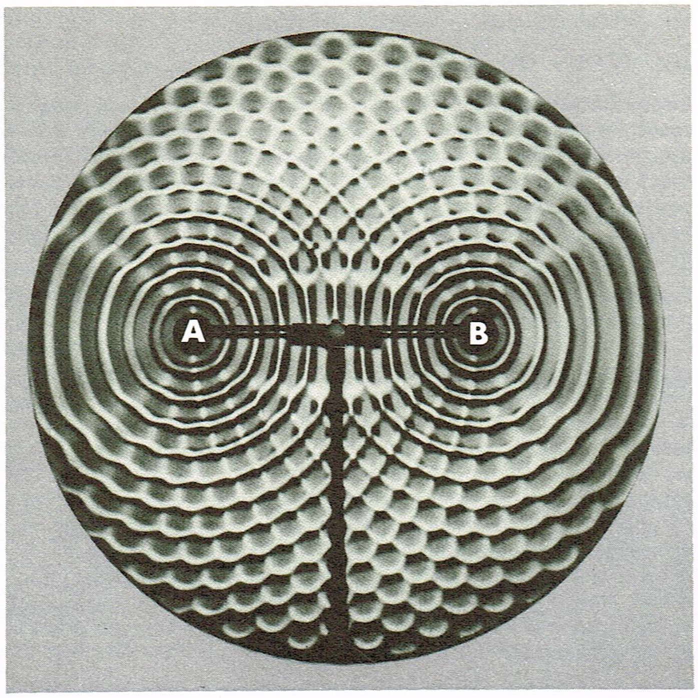


diatomées

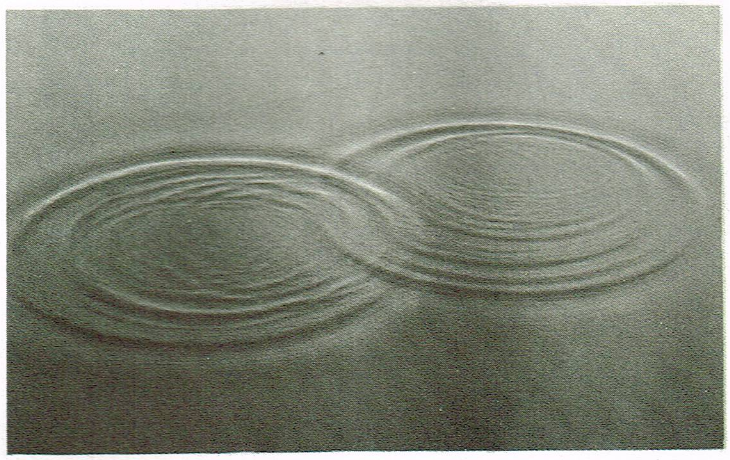




coupe de feuille d'une plante



Phénomène d'interférence produit à la surface du mercure et photographié avec l'ondoscope Charron. A et B sont deux centres d'ébranlement d'où partent deux systèmes d'ondes concentriques engendrées synchroniquement. *Ci-dessous,* autre phénomène d'interférence, mais les ondes ne sont pas en synchronisme.



les interférences

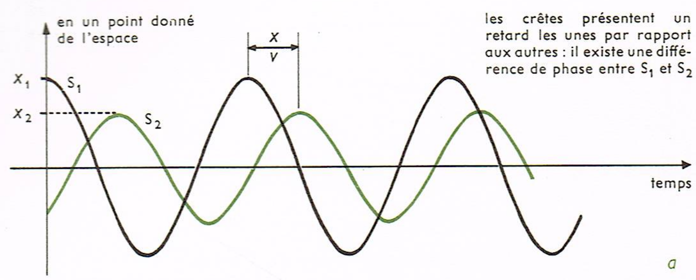
Le modèle des ronds dans l'eau, celui des ondes mécaniques, reste toujours commode. Une pierre tombe : un système d'ondes s'établit. Qu'une seconde pierre tombe un peu plus loin : un second système s'établit à son tour et se combine avec le premier. Certains maximums semblent se renforcer entre eux; en d'autres points de la surface, des maximums de l'un des trains d'ondes se trouvent compensés par des minimums du second train. On dit qu'il y a interférence, addition vectorielle des deux mouvements ondulatoires.

En optique, il en est de même : la superposition de deux trains d'ondes de même fréquence est susceptible de donner lieu à un phénomène d'interférence.

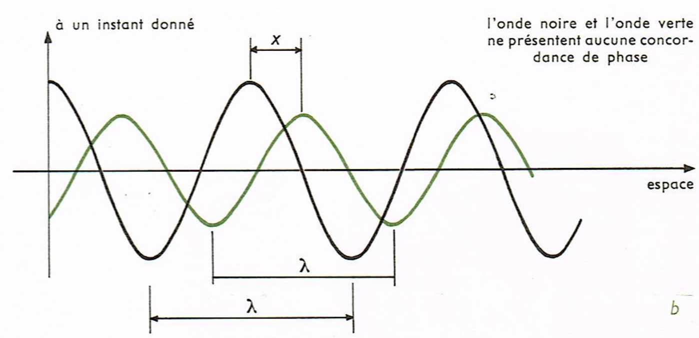
Que le mot « susceptible », jetant une note dubitative dans cette affirmation, soit maintenant bien vite expliqué!

Tout émetteur élémentaire de lumière se comporte comme un doublet rayonnant, fonctionnant indépendamment de ses voisins. Le champ électrique peut être, à un instant donné, déterminé par une amplitude et une orientation quelconques (dans le plan perpendiculaire au rayon lumineux). Dans ces conditions, comment espérer que les mouvements d'une multitude d'émetteurs-doublets constituant la source lumineuse proprement dite et ne présentant aucune relation entre eux puissent s'arranger en un mouvement d'ensemble aussi bien ordonné que celui des ondes à la surface de la cuve? En fait, ils ne s'arrangent pas aussi facilement. Le plus souvent la confusion règne. Il y a **incohérence.**

Comment s'acheminer vers un état plus ordonné dans lequel la superposition des vibrations élémentaires représenterait quelque chose de plus cohérent? On peut, certes, effectuer un triage des longueurs d'onde au moyen d'un filtre (1) isolant une seule radiation. Mais cela ne suffit pas, puisqu'il faut encore que les vibrations soient parallèles; or, ce n'est pas possible



Représentation de deux ondes de même période en fonction du temps (*a*) et en fonction de la distance à une origine déterminée (*b*).

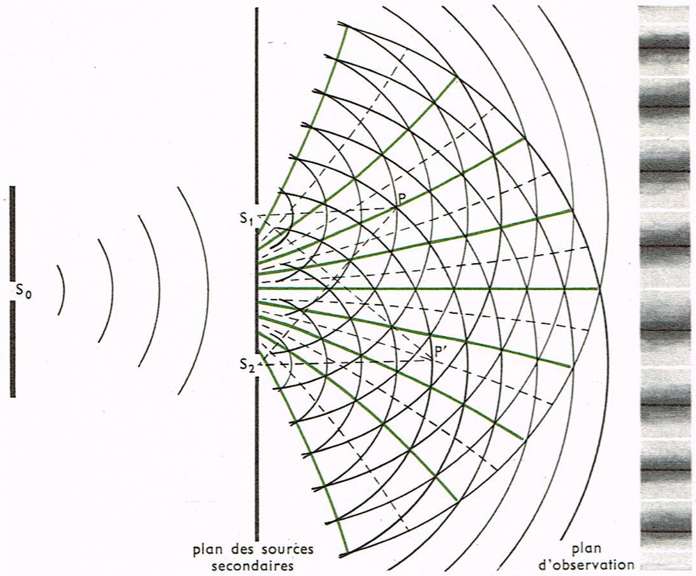


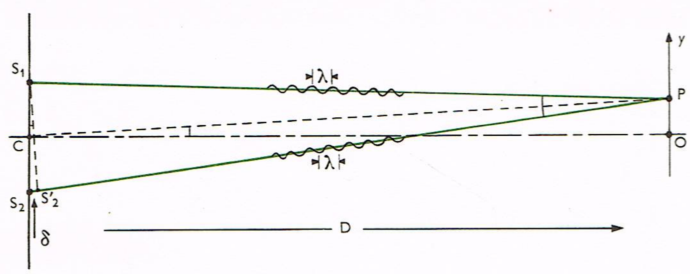
avec la lumière naturelle (2). On tentera de ne mettre en présence que des sources qui vibrent de la même manière au cours du temps. **Il est beaucoup plus simple de superposer deux vibrations provenant d'un même point source.** En tout point de l'espace, les phases de ces deux vibrations sont liées, l'avance ou le retard des deux ondes correspondantes sont parfaitement définis. On dira qu'il y a **cohérence.**

(1) Quelle que soit la nature de ce filtre (verre coloré, prisme dispersif permettant de séparer les couleurs, etc.), il ne permettra pas d'isoler une onde de période unique, mais une « bande » suffisamment étroite pour que l'on puisse parler de lumière à peu près monochromatique.

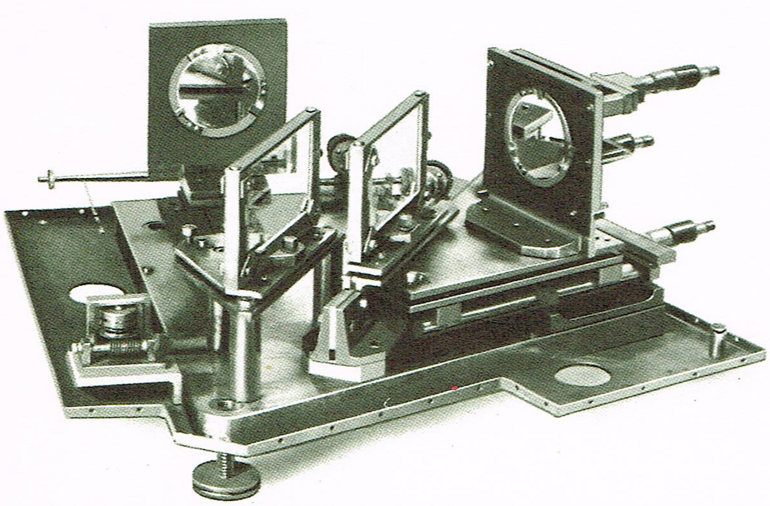
(2) Des ondes polarisées, dont on peut choisir l'orientation de manière à aboutir aussi à des interférences, sont réalisables au moyen du Polaroïd ou de substances biréfringentes.

Phot. Larousse.





Expérience des trous de Young.
Interférences produites par deux sources secondaires S_1 et S_2 éclairées par une même source initiale S_0. La superposition de deux maximaux (par ex., S_1P et S_2P) produit un maximum : les deux ondes s'ajoutent et, dans le plan de l'observateur, donnent une frange claire. En revanche, là où un maximum se superpose à un minimum (par ex., S_1P' et S_2P'), les ondes s'annulent et on observe une frange sombre.



Interféromètre de Michelson.

une expérience fondamentale : les trous de Young

La manière la plus simple de concevoir et de réaliser pratiquement les conditions de cohérence est de placer deux trous devant une source monochromatique ponctuelle. Les amplitudes et les phases relatives de ces sources « secondaires » restent ainsi constamment liées et ne peuvent varier l'une par rapport à l'autre de façon aléatoire. L'addition géométrique des mouvements produira le phénomène d'interférence.

Soit donc S_1 et S_2 deux trous percés dans un écran et éclairés par une source S_0. Les vibrations issues de S_0 atteignent S_1 et S_2, qui se comportent à leur tour comme des sources synchrones rayonnant, au-delà de l'écran, dans toutes les directions de l'espace. L'analogie avec les rides à la surface de l'eau est ici complète : la superposition de deux maximums produit un maximum, tandis que là où un maximum se superpose à un minimum, ils s'annulent mutuellement. La figure permet de suivre les lignes de maximums et de minimums du mouvement résultant. Dans un plan perpendiculaire au plan de la figure, et suffisamment éloigné pour qu'on puisse y considérer les ondes comme planes, on observe que les maximums et les minimums sont équidistants; les lignes alternativement lumineuses et sombres sont des franges d'interférence dont l'intensité varie sinusoïdalement.

Le schéma de la figure ci-contre met en évidence quelques notions nécessaires pour la suite. P étant un point quelconque du plan d'observation, son état d'interférence est déterminé par les situations respectives en ce point des ondes venant de S_1 et S_2. Il y a maximum en P lorsque les deux vibrations sont en phase, c'est-à-dire encore lorsque les trajets S_1P et S_2P diffèrent d'un nombre entier de longueurs d'onde.

On peut ainsi s'attendre à observer, peut-être même directement, un phénomène correspondant à une distance de l'ordre de la longueur d'onde de la lumière. Les vibrations arrivant sur l'axe en phase (par raison de symétrie), l'interfrange *i* sera donnée par la position du premier maximum, c'est-à-dire par celle du point P correspondant à une différence de marche $\delta = S_2P - S_1P = \lambda$. L'égalité des angles $S_2S_1S'_2$ et OCP donne, pour valeur de l'interfrange : $i = \frac{\lambda D}{l}$. Soit pour deux trous séparés de 1 mm et éclairés par une source bleu-vert ($\lambda = 0{,}5\mu$), le plan d'observation étant distant de 1 m (les calculs sont faits en millimètres) : $i = \frac{(0{,}5 . 10^{-3})\ (1\ 000)}{1} = 0{,}5$ mm. Quelques applications de ce phénomène seront données un peu plus loin, qui mettront en œuvre ce puissant moyen d'exploration et de traduction fourni par les interférences dans la course vers l'infiniment petit.

la diffraction

Essayons maintenant d'expliquer la limite des possibilités du microscope dans la visualisation des détails de plus en plus fins. Cette limite, inhérente à la nature physique de la lumière, nous allons la faire ressortir de la théorie de la diffraction fondée sur le principe de Huygens et sur ce qui vient d'être développé des phénomènes d'interférence.

Lorsqu'on regarde une source lumineuse à travers un tout petit trou percé dans une feuille de bristol ou de clinquant, de la lumière s'échappe en dehors du cône géométrique qui s'appuie sur le contour du trou. A travers des rideaux fins, la lampe de la rue semble multipliée, irisée; les couleurs apparaissent d'autant plus nettes qu'on s'éloigne de la région centrale, celle où passent les rayons directs.

Ce sont autant de phénomènes de diffraction : chaque élément du petit trou, chaque maille du rideau se comporte comme une

Phot. Viénot

source lumineuse fonctionnant pour son propre compte — source secondaire qui, à son tour, rayonne dans toutes les directions (principe de Huygens).

Si, dans le bristol de tout à l'heure, on découpe une fente fine et que l'on regarde à travers elle une source éloignée, disons le Soleil, on voit apparaître parallèlement à la découpure des bandes alternativement sombres et brillantes.

Le phénomène est difficile à interpréter géométriquement : il n'y a plus interférence entre les ondes issues de deux points, mais entre les ondes issues d'une infinité de points, ceux de toute la surface de la fente. Aussi n'en tenterons-nous pas l'analyse, mais nous contenterons-nous d'en faire comprendre les résultats, obtenus mathématiquement.

Pour cela, nous allons étudier d'abord un dispositif de complexité intermédiaire.

les réseaux

Une onde plane arrive sur un écran percé d'un grand nombre de petites ouvertures disposées régulièrement, par exemple une succession de fentes fines parallèles distantes de p (période) ou encore sur des facettes taillées dans un matériau qui pourrait être du verre ou de l'acier poli. Qu'arrive-t-il?

Chaque élément diffracte la lumière (principe de Huygens), et l'on voit que dans toutes les directions définies par $\sin\theta = K\frac{\lambda}{p}$ (K étant entier : ..., -3, -2, -1, 0, 1, 2, 3, ... est appelé « ordre du spectre »), les ondes sont en phase, ce qui correspond à des concentrations de lumière que l'on peut recueillir dans le plan focal de la lentille L et qui représentent les différents ordres du spectre du réseau. Si la lumière incidente comporte plusieurs longueurs d'onde (lumière polychromatique), chaque longueur d'onde λ fournit son propre spectre (aux différents ordres). C'est l'irisation observée à travers un rideau : dans chaque ordre, le rouge est plus dévié que le bleu (λ rouge plus grand que λ bleu). Le réseau est donc un appareil dispersif. Il va jouer le rôle de séparateur de longueurs d'onde (monochromateur) et d'analyseur de lumière (spectrographes et spectromètres modernes). Alors comment ne pas réaliser son importance en analyse spectrale, dans la détermination qualitative et quantitative des éléments d'alliage aussi bien que dans l'étude de la composition du Soleil et des étoiles?

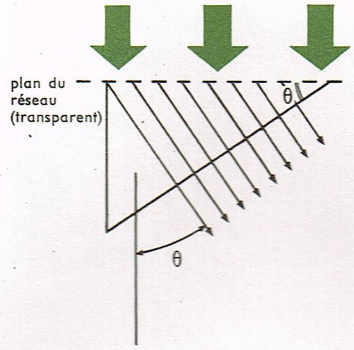
Le pouvoir dispersif en est d'autant plus grand que K est élevé (mais on est limité dans cette direction, les intensités décroissant vite avec l'ordre; de plus, K est limité par la condition $\sin\theta < 1$), que le nombre de traits du réseau est grand (il n'est guère possible de tracer un nombre illimité de lignes suffisamment régulières) et que la période p du réseau est petite. Néanmoins, on sait construire des réseaux obtenus par réflexion et portant plusieurs milliers de traits par millimètre. On se doute que leur réalisation pose de sérieux problèmes techniques dont se ressent leur prix de revient (pour donner un ordre de grandeur, un réseau de 15 cm sur 20, de bonne qualité courante, comportant 1 200 traits par millimètre, coûte plus de 25 000 francs 1963).

phénomènes de diffraction

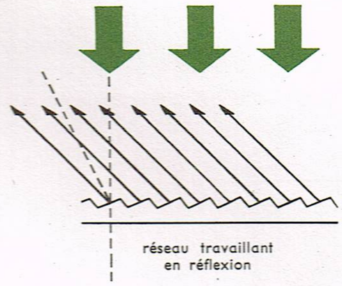
Examinons la figure de diffraction d'une fente éclairée par de la lumière monochromatique : le maximum central, deux fois plus large que les maximums latéraux, est très intense. Les minimums sont définis par la relation angulaire $\sin\theta = \frac{K\lambda}{l}$.

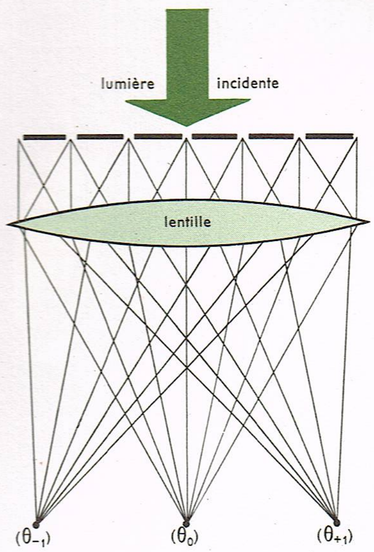
Que peut-on tirer de l'examen de ce phénomène?

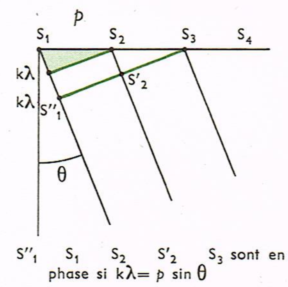
Première constatation : l'intervalle entre les maximums — ou entre les minimums — ainsi que la largeur du maximum central sont inversement proportionnels à la largeur de la fente. Seconde constatation : le passage de la lumière à travers une

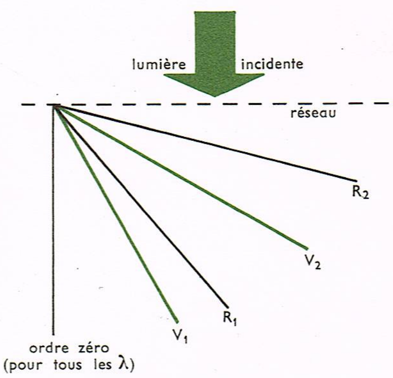


Principe de fonctionnement d'un réseau.









Un réseau décompose la lumière polychromatique en ses différentes couleurs (λ) constituantes.

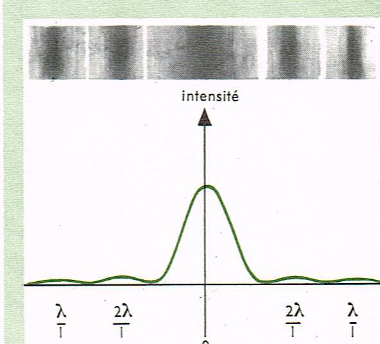
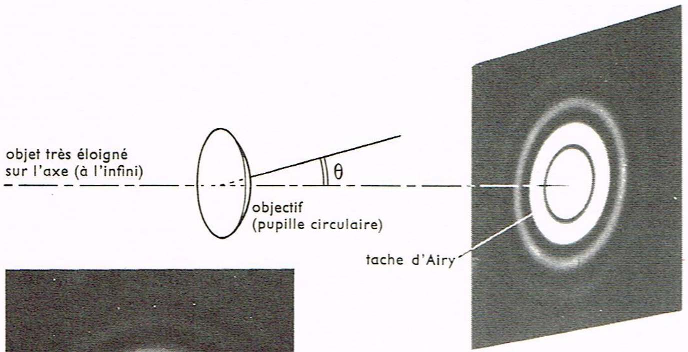
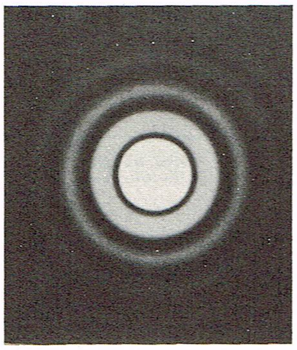
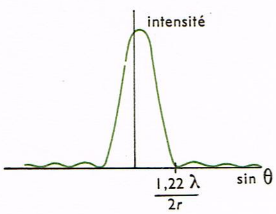


Figure de diffraction d'une fente et graphique représentant l'intensité suivant une ligne parallèle aux franges, c'est-à-dire en fonction de la distance au centre de la tache de diffraction.



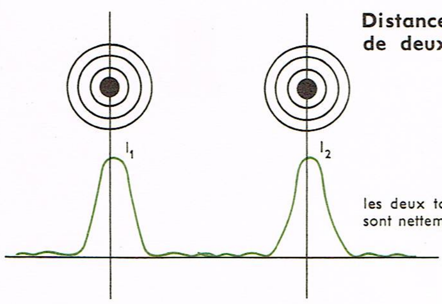


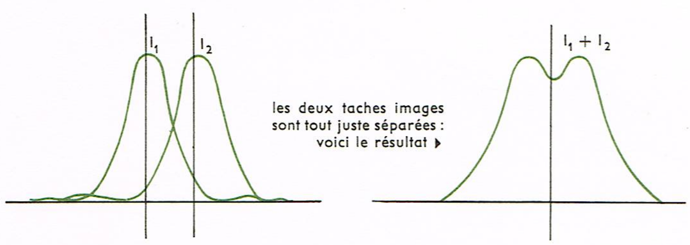
Tache d'Airy.
Ci-dessus,
formation de la tache;
Ci-contre,
photographie des anneaux de diffraction.

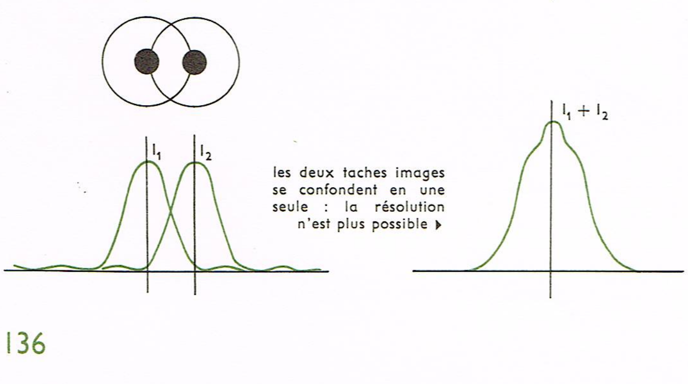


Répartition d'intensité suivant une méridienne.

Distance minimale des centres de deux taches de diffraction.

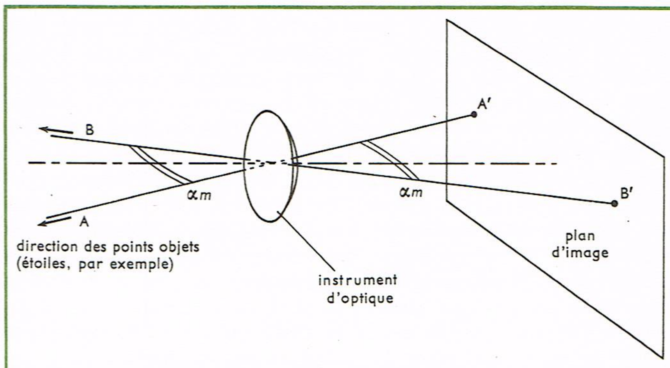






fente, ou, de façon générale, une pupille quelconque de dimension finie, donne lieu à un phénomène de diffraction, qui étale le faisceau incident. Cette conséquence directe de la nature ondulatoire de la lumière est inévitable. Alors, en dehors de toute considération d'aberration (en optique instrumentale, on sait réduire les aberrations géométriques presque autant que l'on veut, à condition d'y mettre le prix), un instrument d'optique est-il capable de donner d'un point une image ponctuelle? **Non, la diffraction l'interdit.** On dira qu'elle « limite la résolution de l'instrument ». A chacun des points d'un objet l'appareil fait correspondre une tache de diffraction — étalement plus ou moins diffus qui empêchera de distinguer un détail de son voisin. Cette tache de diffraction sera liée à la pupille de l'instrument, dont l'ouverture imposera finalement la limite de résolution, le pouvoir séparateur de l'instrument.

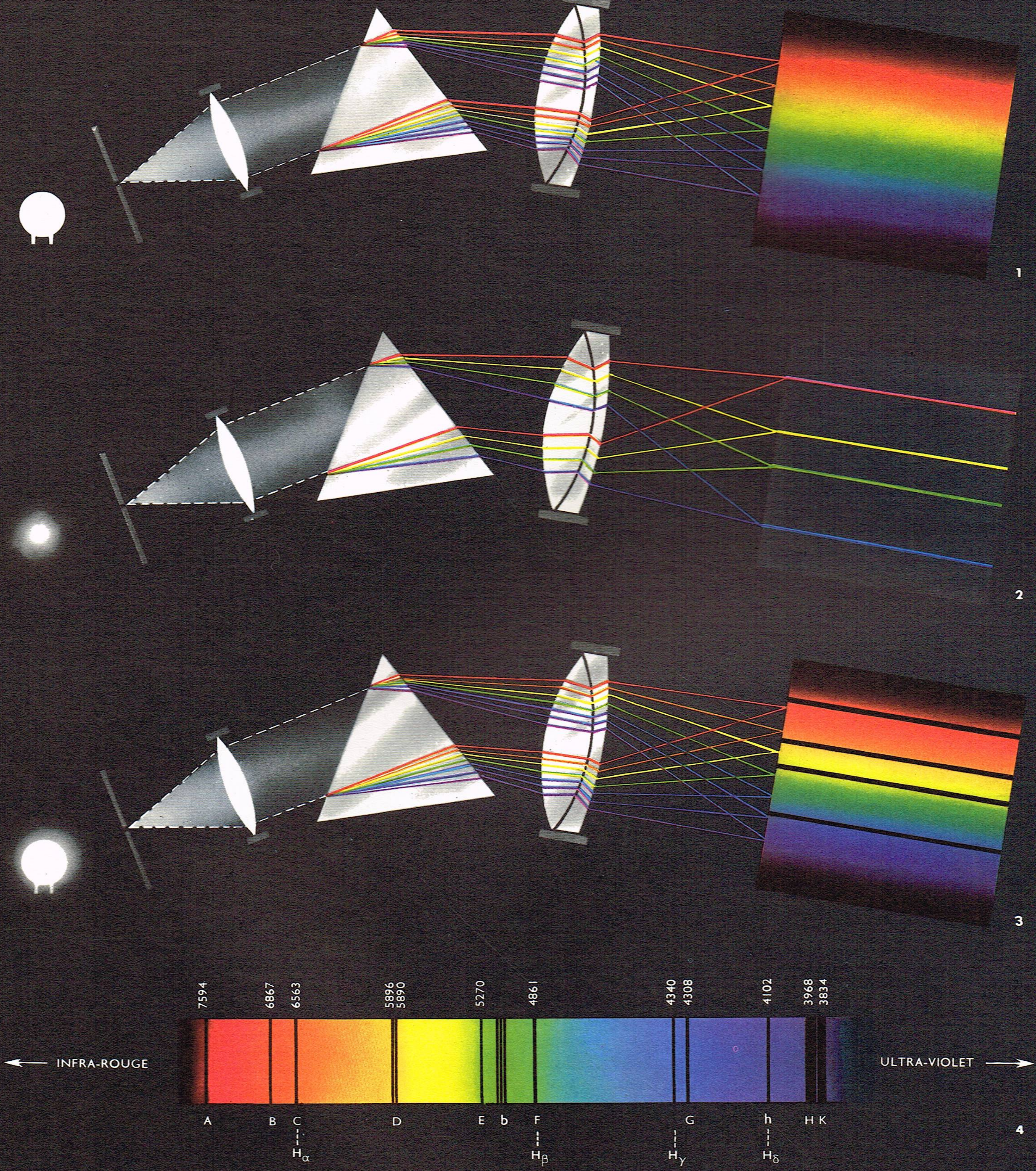
La figure de diffraction d'une pupille aussi simple que la fente est décrite aisément, comme on vient de le voir. La pupille d'un microscope est circulaire, et la détermination de sa figure de diffraction serait à peine moins immédiate. Cette figure de diffraction n'est pas autre chose que l'image d'un point source donnée par l'appareil. On l'appelle la « tache d'Airy ». La section par un plan méridien montre les variations d'intensité, ses maximums et ses minimums. La position du premier minimum donne le rayon de la tache centrale de diffraction en fonction du rayon r de la pupille; on trouve la valeur $\alpha_1 = \frac{1,22\lambda}{2r}$ pour la demi-largeur angulaire du maximum central.



Lorsqu'on les examine à travers un instrument, deux points peuvent être « séparés » si leurs taches de diffraction peuvent être perçues séparément.

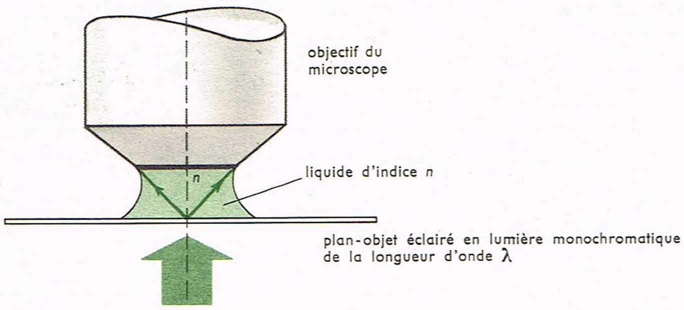
pouvoir séparateur et critère de Rayleigh

Pratiquement, les considérations précédentes amènent à se demander comment fixer quantitativement la capacité qu'a un instrument de distinguer deux points voisins dans l'espace objet. Cela revient à poser la question suivante : quelle est la distance minimale des centres de deux taches de diffraction, jusqu'à laquelle il est possible d'affirmer que ces taches correspondent à deux points objets distincts? La considération de la courbe d'intensité nous fournit la réponse, et puisqu'il faut se fixer une frontière entre la possibilité et l'impossibilité de « résoudre » ces images, on adopte généralement une règle, celle dite « de Rayleigh » : les taches de diffraction sont discernables lorsqu'elles sont séparées par la demi-largeur du maximum central. La distance angulaire minimale de deux points séparables est donc, pour une pupille d'instrument circulaire (voir paragraphe précédent), $\alpha_{min} = \frac{1,22\lambda}{2r}$. Cette valeur minimale est d'autant plus faible que le diamètre $2r$ de la pupille est important. Ainsi le diamètre d'ouverture d'une

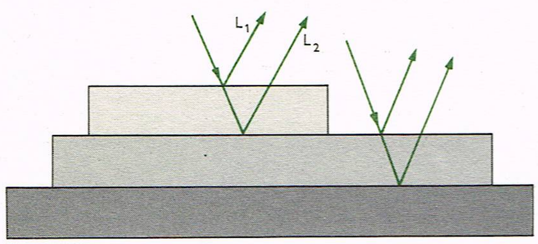


PRINCIPE DE L'ANALYSE SPECTRALE

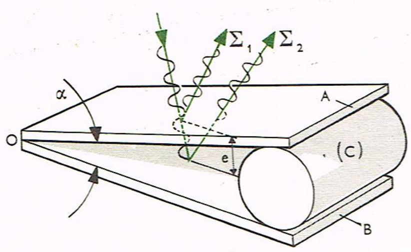
1. Décomposition et dispersion de la lumière blanche par un prisme. Spectre continu fourni par une source incandescente (lampe électrique, photosphère solaire...); 2. Spectre discontinu, composé de raies d'émission, fourni par un gaz luminescent (lampe à vapeur de mercure, chromosphère solaire...); 3. Spectre discontinu, coupé de raies d'absorption, fourni par une source incandescente entourée d'une couche de gaz absorbant (soleil : photosphère et couche renversante); 4. Principales raies d'absorption du spectre solaire (raies de Fraunhofer).



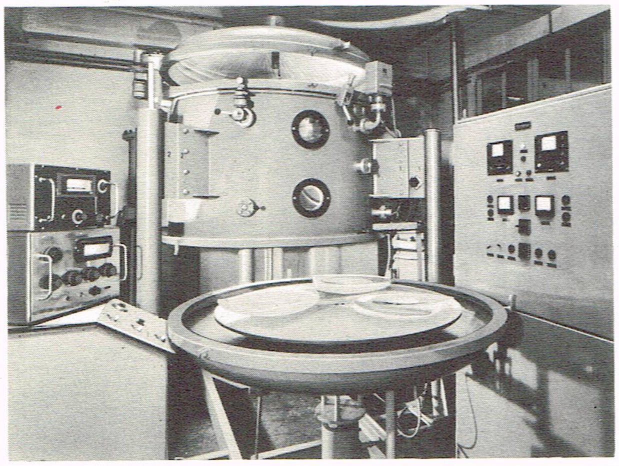
Faisceau de lumière à l'entrée de l'objectif du microscope : l'angle d'ouverture dépend du liquide d'immersion employé.



Un rayon incident se sépare en deux parties lorsqu'il rencontre la première surface : il y a réflexion partielle et réfraction, puis une réflexion sur la seconde « pile ». Les rayons tels que L_1 et L_2, provenant d'un même rayon incident, vont interférer.



Le coin d'air : les rayons Σ_1 et Σ_2 font entre eux un angle, le même que celui des deux lames.



Appareil d'évaporation sous vide, pour déposer des couches antireflets à la surface des lentilles.

lunette astronomique doit être aussi grand que possible si l'on veut « séparer » des étoiles proches sur la sphère céleste. Le pouvoir séparateur théorique de la lunette de Yerkes (de diamètre 102 cm) serait (pour une longueur d'onde moyenne de 0,5 μ, soit 5.10^{-4} mm), $\alpha_{min} = \frac{1{,}22\,(5.10^{-4})}{1\,020}$ environ, soit de 0,1 seconde d'arc en gros, c'est-à-dire qu'elle serait capable — théoriquement, bien entendu — de distinguer deux points distants d'un mètre environ situés à 1 500 km!

Et notre œil, que peut-il séparer? Son maximum de sensibilité se trouve dans le jaune ($\lambda = 0{,}58.10^{-3}$ mm). Lorsque l'ouverture de la pupille est de l'ordre de 3 mm ($2r = 3$ mm), la limite de résolution est de $\frac{1{,}22\,(0{,}58.10^{-3})}{3} = 2{,}4.10^{-4}$ (soit une minute d'arc). Habituellement, on considère que l'œil peut distinguer des détails voisins distants (angulairement) d'une minute environ (1 mètre à 5 km).

Le pouvoir séparateur du microscope dépend de l'indice n du milieu objet et de l'inclinaison u des rayons extrêmes. La distance minimale d_m de séparation de deux points du plan objet est donnée par la relation $\mathbf{d_m = \frac{1{,}22\,\lambda}{2n \sin u}}$. La quantité $n \sin u$ est appelée « ouverture numérique de l'objectif ». Cette relation montre l'intérêt qu'il y a de travailler avec une longueur d'onde assez faible (en lumière bleue, par exemple, mais l'acuité visuelle est alors moindre), et surtout d'augmenter l'ouverture numérique. Pour un objectif bien corrigé, une valeur de sin u voisine de 0,92 est déjà très honnête. La seule possibilité est d'agir sur n, c'est-à-dire de trouver un liquide possédant un fort indice de réfraction dans lequel sera immergé l'objet à examiner. Certaines huiles ($n = 1{,}92$) permettent ainsi de descendre à une limite d_m de l'ordre de 0,10 à 0,15 μ. Ainsi le grossissement optimal se trouve du même coup déterminé, et il ne serait d'aucune utilité d'augmenter la puissance très au-delà de celle qui est nécessaire à la vision d'un détail de cette dimension.

Nous touchons ainsi à l'une des limites du microscope optique. Un pouvoir séparateur, une limite de résolution beaucoup plus importante, c'est-à-dire en fait une possibilité plus grande de discerner les détails, seront fournis par le microscope électronique, qui permet (λ très petit) de descendre à une valeur d_m de l'ordre de l'angström ou au-dessous, comme nous l'avons déjà vu au tome I).

lames minces : franges d'égale épaisseur

Une goutte d'huile sur l'eau s'étale et produit des jeux de lumière de colorations parfois surprenantes. Il s'agit d'un phénomène d'interférence entre des rayons réfléchis sur deux couches successives — deux empilements de molécules d'acide gras, par exemple —, dont l'épaisseur peut être de l'ordre de la centaine d'angströms. Lors de l'interférence de rayons tels que L_1 et L_2, il y a formation de franges caractéristiques. Si la lumière incidente n'est plus chromatique, chaque longueur d'onde produira son propre système de franges, d'où l'absence ou la prépondérance, dans la lumière réfléchie, du bleu, du jaune, du violacé, etc.

Une expérience très simple peut être réalisée au moyen de deux lames de verre entre lesquelles une cale mince (C) laisse subsister un coin d'air d'angle faible, de quelques minutes d'arc seulement (e de l'ordre de λ). La lumière tombant sur ce coin se réfléchit partiellement sur chacune des faces OA et OB. Dans une région où l'épaisseur est e, on peut définir l'état d'interférence comme précédemment. Lorsque l'épaisseur du

Phot. A. Papillon.

parcours du rayon dans le coin d'air varie de $\lambda/2$, la différence de marche varie de λ, et l'on se retrouve dans un état analogue. Des variations d'épaisseur très faibles (une fraction de λ) sont ainsi mises en évidence.

lames épaisses : franges d'égale inclinaison

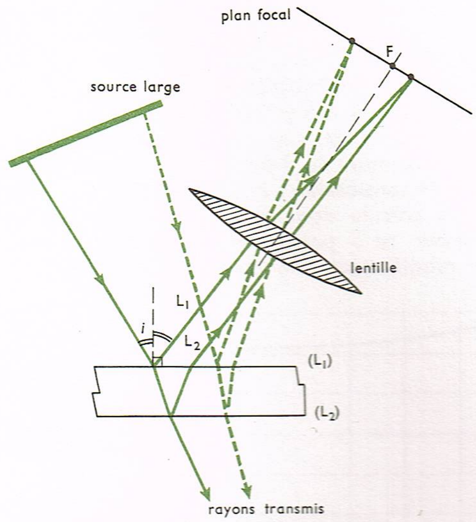
Un rayon lumineux tombe sur une lame à faces rigoureusement parallèles. Il y a réflexion sur la face avant (L_1), tandis que la lumière qui a traversé cette face se réfléchit à nouveau, en partie, sur la face arrière (L_2). Les rayons L_1 et L_2 émergent parallèlement (de multiples réflexions se produisent, mais la deuxième et les suivantes sont suffisamment affaiblies pour ne plus intervenir). Ainsi, chaque rayon provenant de la source est susceptible de la même histoire et chaque couple L_1 L_2 interfère à l'infini, ou encore dans le plan focal d'une lentille disposée à cet effet. La différence de marche suivant des rayons tels que L_1 et L_2 n'est fonction que de l'angle *i*. Donc tous les rayons de même inclinaison interfèrent à la même distance de F. Le résultat est un ensemble de franges circulaires : les **anneaux à l'infini,** ou franges d'égale inclinaison. Le profil photométrique n'est plus tout à fait sinusoïdal. On peut montrer que les maximums se succèdent suivant des rayons qui croissent comme les racines carrées des entiers successifs, selon la formule $R_2 = \frac{K\lambda}{e}$.

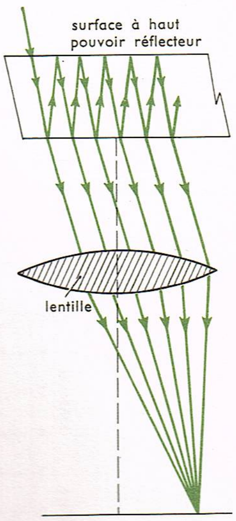
Lorsque la lame L est un verre dont les surfaces planes n'ont subi aucun traitement, seuls les rayons L_1 et L_2 doivent être pris en considération : ils transportent pratiquement toute l'énergie véhiculée par le rayon initial (à l'absorption près). Si, maintenant, les faces de L sont recouvertes d'une couche transparente à haut pouvoir réflecteur, un système de réflexions s'établit à l'intérieur de la lame : un même incident fournira de nombreux émergents, dont l'intensité s'affaiblira peu lorsqu'on passera de l'un d'eux au suivant (on travaille alors en transmission). Le phénomène aura la même allure que le précédent, mais les anneaux seront plus lumineux et plus fins. Pratiquement, la lame L sera constituée par deux lames de verre dont les faces en regard, très planes, auront été traitées : c'est l'étalon de Fabry-Pérot. La théorie conduit encore à la même relation $R^2 = \frac{K\lambda}{e}$. Chaque longueur d'onde donnant lieu à un système d'anneaux, l'étude du profil de ceux-ci fournira des renseignements sur la composition chromatique de la lumière incidente. La précision extrême des méthodes mettant à profit ce type de dispositif en a fait un puissant moyen de recherche, en particulier dans le domaine de la spectroscopie à haute résolution (analyse spectrale, structure hyperfine, filtres interférentiels).

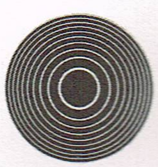
les couches antireflets

Au début de ce siècle, on avait remarqué que certains appareils optiques anciens étaient plus transparents que d'autres plus récents; la cause semblait résider dans une altération superficielle des verres. C'est ce qu'explique le schéma élémentaire que nous avons donné de formation des franges d'interférences dues aux couches minces. Pour une inclinaison moyenne donnée du rayon incident, on peut déterminer l'épaisseur e d'une couche transparente d'indice *n* pour laquelle la marche optique entre les rayons 1 et 2 réfléchis sur les deux faces de la couche diffère d'une demi-longueur d'onde $\lambda_0/2$ de la lumière incidente (à un nombre entier de λ_0 près). Dans ces conditions, les vibrations qui correspondent à la longueur d'onde en question λ_0 se détruisent par interférence. Il n'y a pas de lumière réfléchie dans cette longueur d'onde.

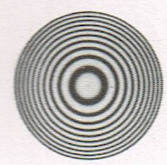
S'il n'y a pas de perte par réflexion, on peut s'attendre que

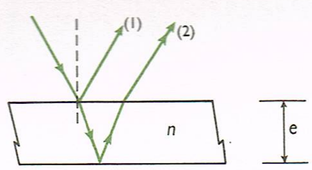




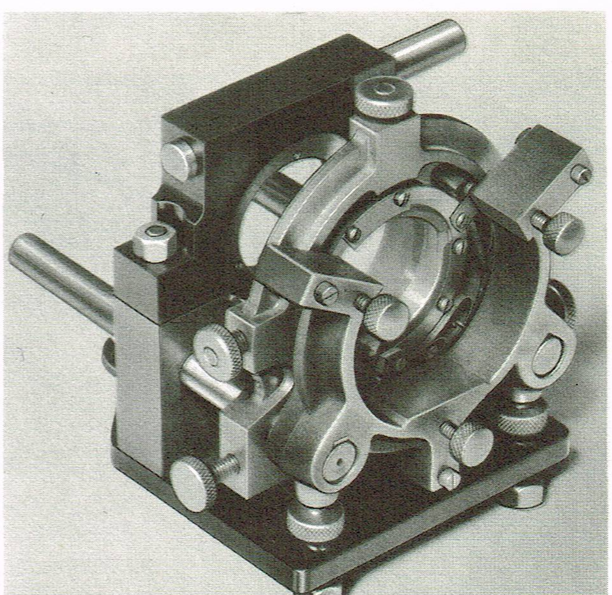


Anneaux obtenus avec des lames : *à gauche*, avec lames non traitées (2 ondes); *à droite*, en réflexions multiples avec lames traitées (étalon de Fabry-Pérot).





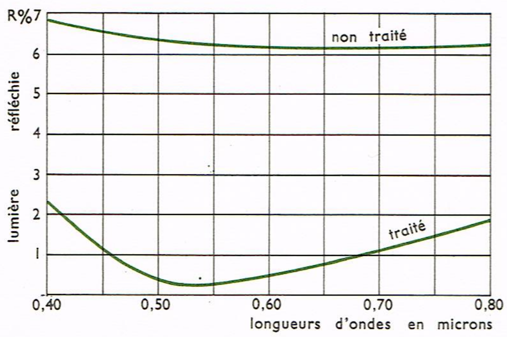
Dédoublement du faisceau incident sur une lame partiellement réfléchissante.



Etalon de Fabry-Pérot.

Phot. Viénot.

toute la lumière soit transmise. Mais attention, le traitement n'est réalisable que pour une seule longueur d'onde, c'est-à-dire une seule couleur. Imaginons que l'on veuille traiter des jumelles par exemple, pour empêcher toutes les réflexions parasites sur les surfaces (4 à 8 p. 100 de la lumière incidente pour un verre non traité), et, par là, tout dédoublement des images, tout en augmentant la fraction de la lumière transmise. Le maximum de sensibilité de l'œil étant situé au voisinage de $\lambda = 0{,}58\ \mu$, on calcule donc la couche mince à déposer sur les objectifs pour ce λ particulier. La lumière transmise ne subit aucune modification (de couleur en particulier). En



Courbes de bandes passantes pour un verre non traité et pour le même verre pourvu d'une simple couche antiréfléchissante (d'après MTO).

regardant la surface de l'objectif par réflexion, λ est absent de la lumière réfléchie qui contient encore les extrémités du spectre visible, d'où une coloration pourpre de cette surface. Pour la même raison, certains objectifs photographiques semblent colorés en jaunâtre : la sensibilité du film photographique étant différente de celle de l'œil, des longueurs d'onde correspondant au bleu ont été ici éliminées.

L'attaque superficielle du verre n'est pas une méthode couramment employée aujourd'hui pour réaliser une couche antireflet! On préfère évaporer la substance transparente (fluorure de magnésium, cryolite) d'indice convenable (en général compris entre 1,3 et 1,4) sous un vide élevé (pression réduite à quelques microns de mercure) dans une cloche qui contient le verre à traiter. Il se condense sur celui-ci une couche dont l'épaisseur uniforme et faible peut être contrôlée précisément par la couleur de la lumière réfléchie pendant la formation du dépôt.

les filtres

Le problème est en quelque sorte réciproque du précédent : il s'agit de concevoir des systèmes ne laissant passer qu'une bande très étroite de longueurs d'onde. On dépose sur un support une série de couches minces d'épaisseur et d'indices appropriés, disposées en sandwichs. Les retards de phase produits à la traversée des couches et lors des réflexions modifient le facteur de transmission de l'ensemble. Ainsi, des largeurs de bandes passantes de quelque 50 angströms, avec un facteur de transmission de 30 p. 100, peuvent être réalisées autour d'une ou plusieurs longueurs d'onde moyennes dans le spectre visible (les transmissions entre deux bandes consécutives étant réduites à 1 p. 1 000 environ).

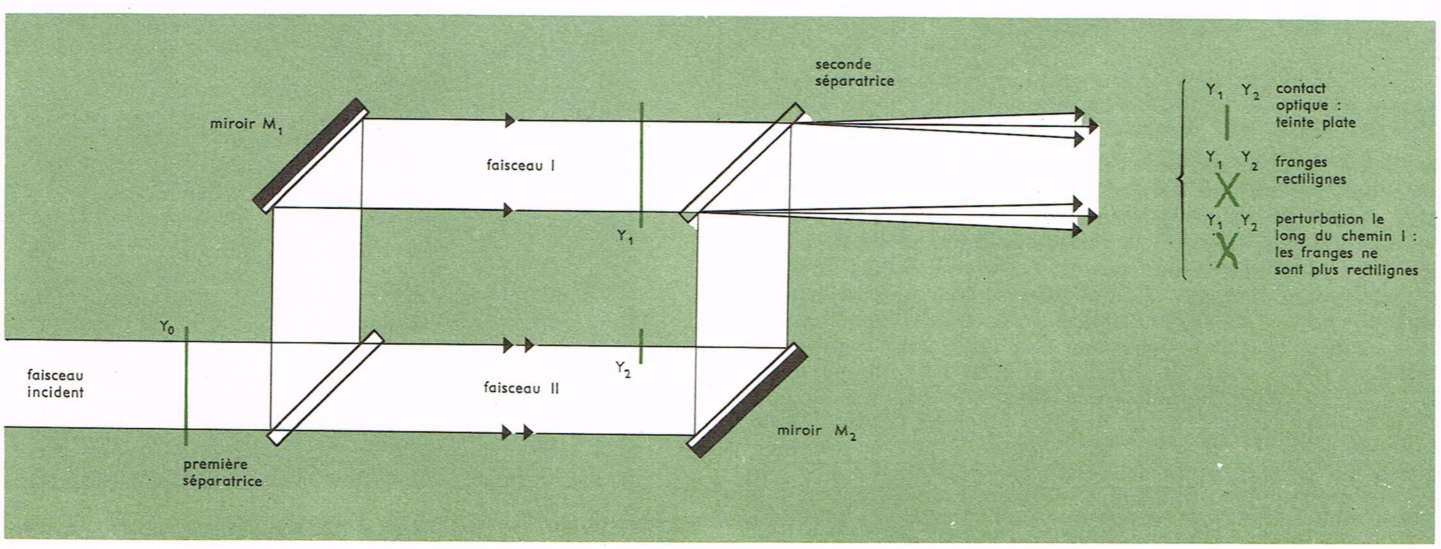
interféromètres et interférométrie

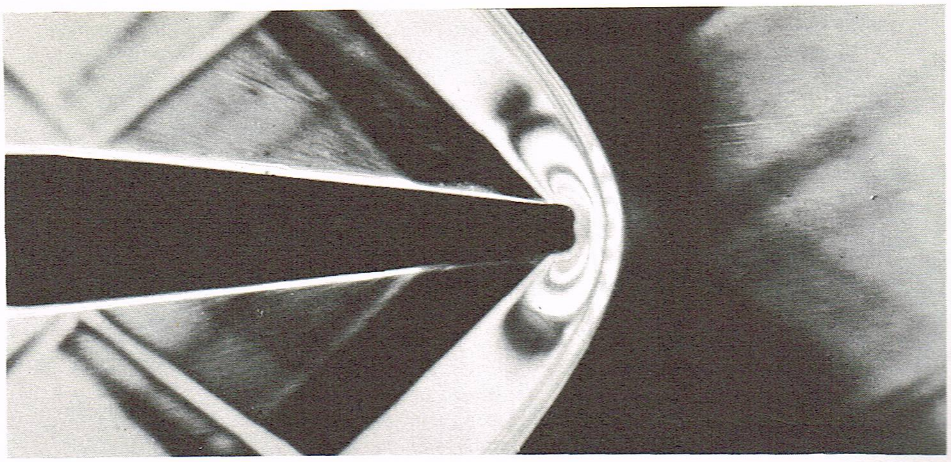
Le processus de dédoublement et recombinaison se réalise encore d'une autre façon : deux faisceaux — ou encore deux ondes — issus d'une même source initiale, qui ont suivi des chemins optiques quelque peu différents, peuvent donner lieu à des phénomènes d'interférence lorsqu'on les regroupe suivant l'un des procédés décrits plus haut.

interféromètre de Mach-Zehnder et avions supersoniques

Imaginons donc l'arrivée d'une onde plane sur une surface semi-transparente, telle qu'une glace légèrement métallisée (lors de la formation de dépôt sous vide d'une couche réfléchissante, on peut contrôler le coefficient de réflexion). Il y a dédoublement en deux ondes, qui chemineront après la traversée de cette glace, appelée « séparatrice », soit vers le miroir M_1, soit vers le miroir M_2.

Principe de l'interféromètre de Mach-Zehnder.





Visualisation par strio-interférométrie de l'écoulement à Mach 1,96 autour du nez d'un profil.



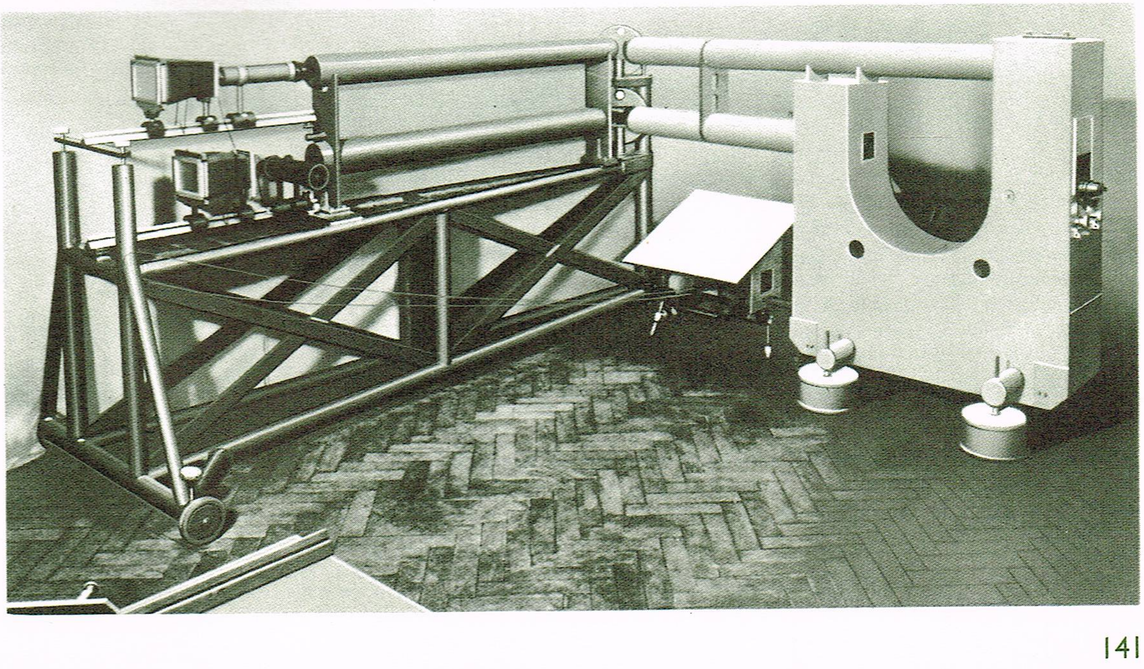
Tube métallique de 1 m de longueur avec bandes d'interférence.



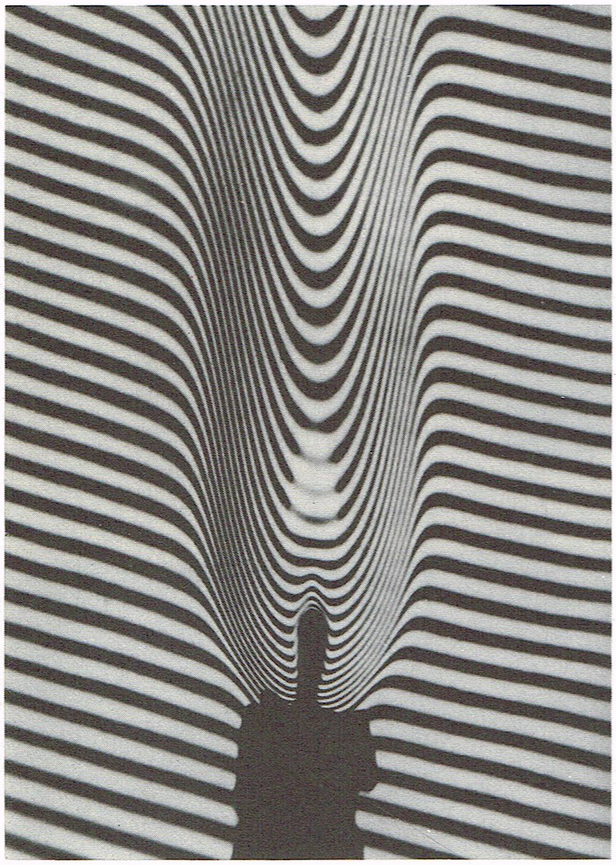
Le même tube chauffé de 1 °C.

Variations de température autour d'un corps chauffé mises en évidence par l'interféromètre de Mach-Zehnder.

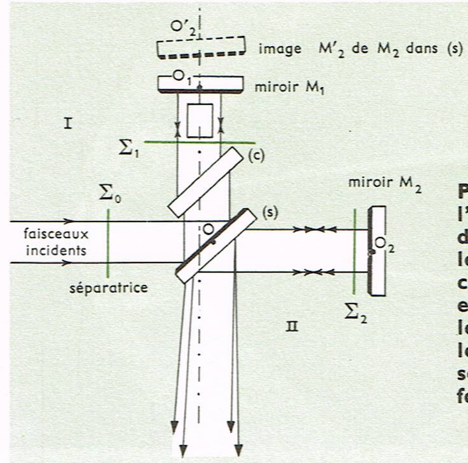
Interféromètre de Mach-Zehnder.



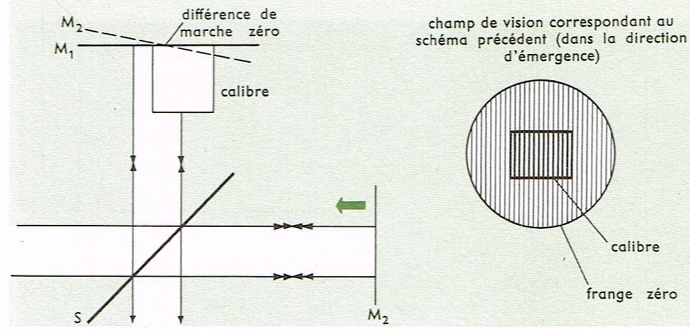
Phot. O. N. E. R. A. et Carl Zeiss.



Courants de convection au-dessus de la flamme d'une bougie vus au Michelson.



Principe de l'interféromètre de Michelson : le rôle de la compensatrice (c) est d'égaliser les trajets I et II, les faisceaux séparés sur la face avant de (S).



Principe de la mesure d'une épaisseur au moyen de l'interféromètre de Michelson.

Les deux faisceaux changent de direction (c'est la deuxième fois pour le faisceau Y_1, et la première pour le faisceau Y_2). La seconde séparatrice permet le regroupement de nos deux ondes. En réglant les éléments optiques, miroirs et séparatrices, on peut fort bien faire coïncider les ondes Y_1, Y_2 à la sortie de l'appareil (1) : c'est ce qu'on appelle le « contact optique ». On peut aussi les incliner d'un petit angle l'une par rapport à l'autre : le système se comporte alors comme un coin d'air.

Un système de franges rectilignes apparaît; son centre, là où la différence de marche est nulle, est situé à l'intersection entre Y_1 et Y_2. Enfin, toute perturbation se produisant le long du chemin 1, par exemple, se traduira par une distorsion des franges, qui suivront en quelque sorte les lignes de niveau. C'est ce qui se passe dans les installations d'étude des maquettes d'avion ou de l'onde de choc due aux projectiles, utilisant les souffleries supersoniques. Une maquette étant placée en travers du trajet 1, une veine gazeuse circule à grande vitesse à l'intérieur d'un tuyau contenant le projectile ou l'aile d'avion réduite, et dans les parois duquel sont aménagées des fenêtres faisant passer le faisceau lumineux. Les remous, les tourbillons provoquent des variations locales de l'indice de l'air autour de la maquette. A ces variations correspondent des déformations locales de Y_1, mises en évidence lorsque cette onde se combine à Y_2 non perturbée. La sensibilité de cette méthode est illustrée dans la série des photos de la page 141, l'agitation produite étant moins grande que dans le cas ci-dessus.

(1) A condition toutefois que les éléments en question ne changent pas la forme de l'onde; on suppose ici leur qualité telle qu'ils n'apportent aucune perturbation.

l'interféromètre de Michelson (1895)

Il tire sa célébrité à la fois des mesures qui ont été faites dès la fin du siècle dernier, lorsqu'on a tenté de déterminer la longueur exacte du fameux mètre du Bureau international des poids et mesures (pavillon de Breteuil, à Sèvres) en fonction d'une longueur d'onde bien définie, non susceptible de vieillir au fil des années, et des tentatives effectuées autour d'une expérience que l'on souhaitait décisive en ce qui concerne la théorie de la relativité (voir chapitre suivant).

Le principe de cet appareil ne diffère pas sensiblement de celui du précédent; seule change la nature du parcours des ondes (chacune décrit un aller et retour), mais la théorie du coin d'air est encore à la base du montage. Les longueurs 00_2 et $00'_1$ étant égales, les rayons émergents SOO_2OL et SOO'_1OL donnent lieu au système de franges correspondant au coin d'air formé par M_1 et son image M'_2 (M_2 n'étant pas perpendiculaire à M_1, M_1 et M'_2 font un petit angle entre eux). Interposons un calibre (bloc d'acier d'épaisseur bien définie aux faces opposées si rigoureusement planes et parallèles que l'on puisse encore parler de coin d'air) sur le parcours OO'_1. La différence de marche qu'il provoque produit, dans la région qu'il occupe, le décalage d'un certain nombre de franges (1). Le passage d'une frange à l'autre correspond à une différence

(1) On observe d'une part les franges dues au coin constitué par M_1 et M'_2 (la face A du calibre étant dans le même plan que M_1), et d'autre part les franges dues au coin M'_2B. Les extrémités A et B du calibre étant parfaitement parallèles, la pente des coins est la même et les phénomènes sont analogues, à une translation près, celle qui correspond précisément à l'épaisseur du calibre.

de marche λ et, puisque les ondes font un aller et retour, à une différence d'épaisseur λ/2. On peut régler M_1 et M_2 de façon à placer la différence de marche « zéro » (frange « zéro ») au centre du champ (dans la région située autour du calibre). Rapprochons alors progressivement M'_2 de S en déplaçant le miroir M_2 parallèlement à lui-même. Les franges, elles aussi, se déplacent dans le champ parallèlement à elles-mêmes, et cela sans que varie l'interfrange. Le comptage du nombre de franges qui défilent entre la position de départ et celle où M'_2 coïncide avec le centre de la seconde face, B, du calibre permet de mesurer le nombre de longueurs d'onde, $\frac{AB}{\lambda}$, que contient l'épaisseur AB.

Voici donc un outil dont la précision extraordinaire (moyennant certaines précautions, on peut aller au-delà du vingtième de frange, c'est-à-dire du centième de micron à peu près) permet des mesures très précises (examen des états de surface, des défauts locaux, etc.). Nous en retiendrons une seule application : la mesure du mètre en longueurs d'onde.

la définition du mètre

Ce type d'expérience fut, en effet, adopté pour comparer le mètre international à toute une série de longueurs d'onde fournies par différentes sources. En fait, on a dû passer par des « intermédiaires ». Ainsi, on a d'abord mesuré un calibre de 0,390 6 mm (près de 650 longueurs d'onde, soit près de 1 300 franges à compter) qui fut ensuite comparé à un deuxième calibre d'épaisseur double du premier, et ainsi de suite; la même opération répétée huit fois aboutit à l'étalon de 10 cm (qui, reporté dix fois, doit donner 1 mètre!).

On a finalement décidé de se rapporter à une longueur d'onde particulière (raie rouge très fine du krypton, de

$$\lambda = 6\,057{,}80211\ \text{Å}),$$

qui devient le nouvel étalon pour l'unité de longueur. Et voilà changée notre bonne vieille définition du mètre. Il va falloir apprendre dès les bancs de l'école communale que :

« Le 14 octobre 1960, à 18 heures, la XIe Conférence générale des poids et mesures, à l'unanimité des trente-deux gouvernements représentés, a voté la résolution suivante :

Résolution 6

La XIe Conférence des poids et mesures, considérant :

— que le prototype international ne définit pas le mètre avec une précision suffisante pour les besoins actuels de la métrologie,

— qu'il est d'autre part désirable d'adopter un étalon naturel et indestructible,

décide :

1° Le mètre est la longueur égale à 1 650 763,73 longueurs d'onde dans le vide de la radiation correspondant à la transition entre les niveaux $2p^{10}$ et $5d^5$ de l'atome de krypton 86;

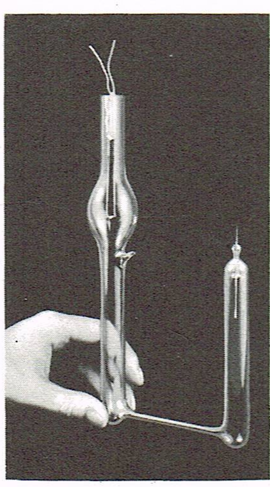
2° La définition du mètre en vigueur depuis 1889, fondée sur le prototype international en platine iridié, est abrogée;

3° Le prototype international du mètre sanctionné par la Ire Conférence générale des poids et mesures de 1889 sera conservé au Bureau international des poids et mesures dans les mêmes conditions que celles qui ont été fixées en 1889. »

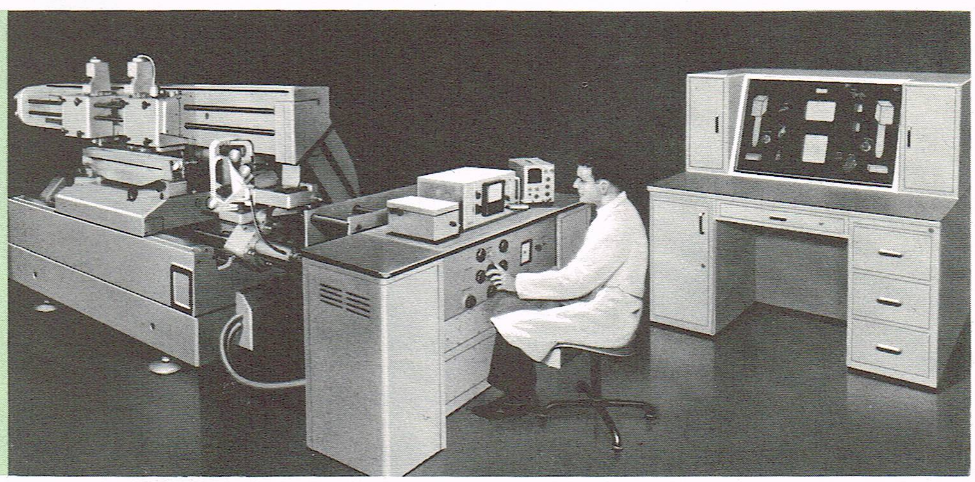
Un décret du 6 mai 1961 fixe le système de mesures obligatoires et définit les unités légales : ainsi disparaît l'ancienne définition du mètre, fondée depuis soixante et onze ans sur un prototype en platine, changée et rattachée maintenant à une longueur d'onde.

Voici donc bouleversée une notion aussi bien établie que celle de la définition du mètre. Mais ce n'est rien devant la révision déchirante des notions d'espace et de temps que nous sommes obligés de faire après une réflexion assez simple sur la vitesse de la lumière, ou devant la nécessité de concilier les données de la mécanique et de l'électromagnétisme, entre lesquels un désaccord choquant se révélera.

Tel sera l'objet du chapitre suivant.

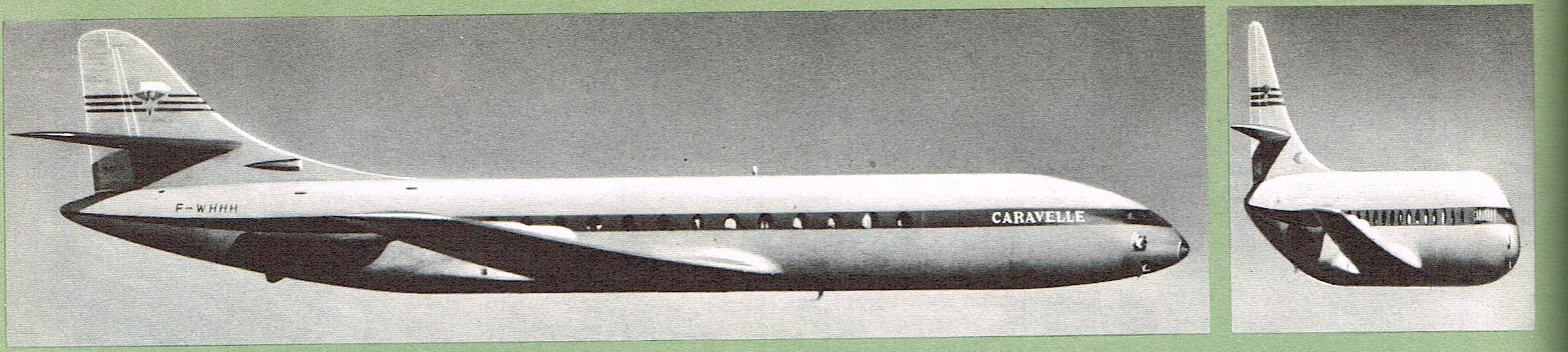


◀ Nouvel étalon optique du mètre.

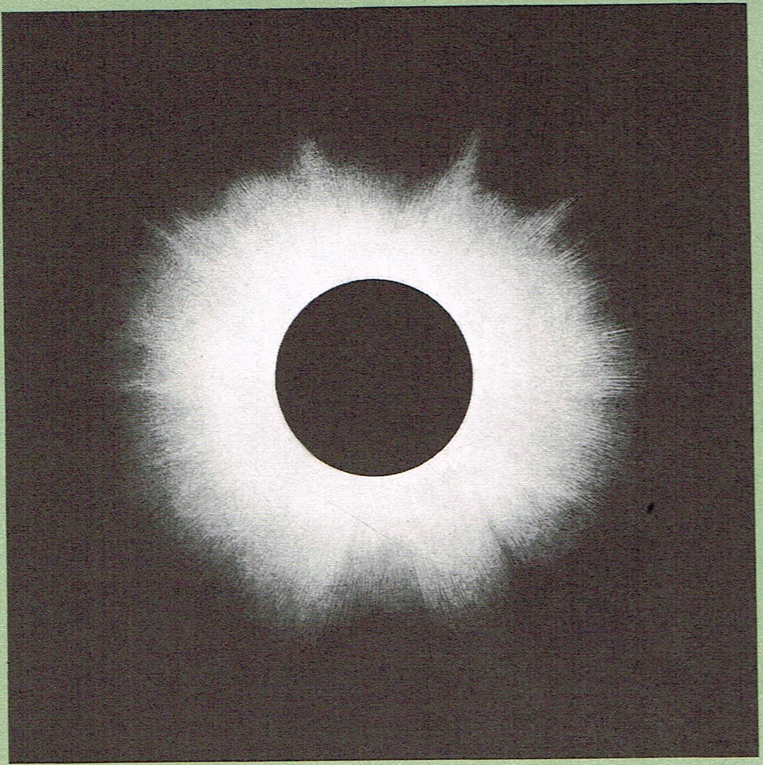


▶ Étalonnage d'un mètre métallique.

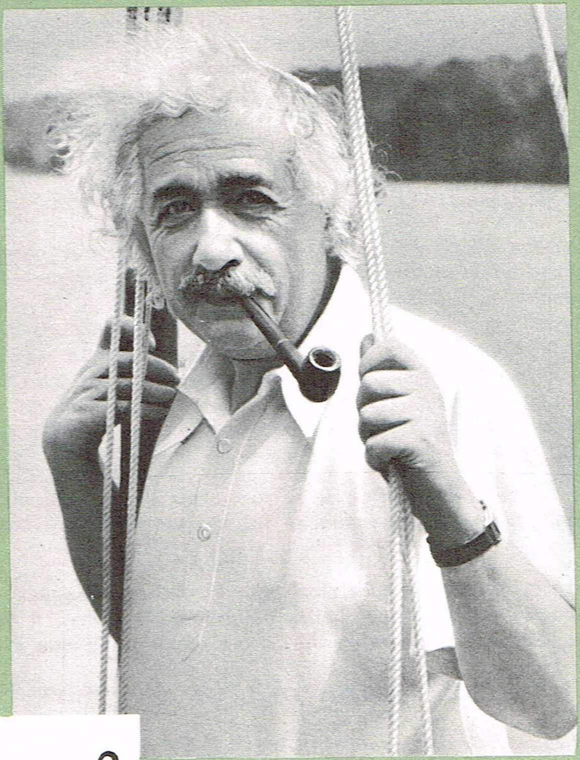
Phot. Bureau international des poids et mesures, Société genevoise d'instruments de physique.



« Caravelle » telle qu'elle est, *à gauche*, et telle qu'on la verrait, *à droite*, si la vitesse de la lumière était de l'ordre de celle de l'avion.

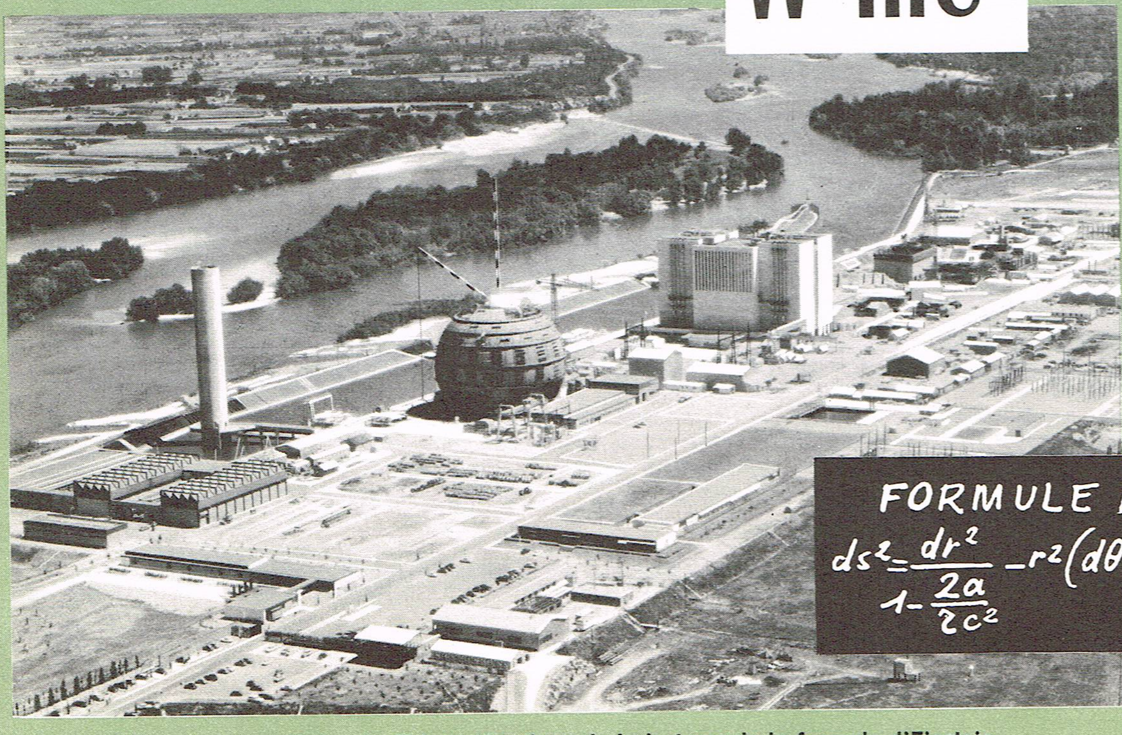


Le Soleil, gigantesque usine atomique qui transforme chaque seconde 4 millions de tonnes de matière en énergie.



Albert Einstein (1879-1955).

$$W = mc^2$$



FORMULE DE SCHWARZCHILD

$$ds^2 = \frac{dr^2}{1-\frac{2a}{rc^2}} - r^2\left(d\theta^2 + \sin^2\theta\, d\varphi^2\right) + \left(1-\frac{2a}{rc^2}\right)c^2 dt^2$$

Phot. Keystone, Larousse, Associated Press, Photothèque française.

La centrale nucléaire de Chinon témoigne de la justesse de la formule d'Einstein.

LA RELATIVITÉ

par

Louis MARIOT

espace et temps ne sont plus absolus

Qu'est-ce que la relativité? Une théorie fantaisiste échafaudée par des originaux maniant les paradoxes et se complaisant à nier des notions admises par tous les hommes de bon sens? ou un monument mathématique édifié à coup de formules transcendantes que seuls peuvent comprendre de rares initiés?

Un coup d'œil rapide et superficiel sur cette théorie appuie la première hypothèse. Qu'on en juge : d'un vaisseau cosmique voguant à 200 000 km/s catapultons un obus à la vitesse de 200 000 km/s; les relativistes diront : la vitesse résultante de l'obus est inférieure à 200 000 + 200 000 = 400 000 km/s (elle vaut exactement 277 000 km/s si l'obus est lancé vers l'avant). Alors 2 et 2 ne font pas 4?

Autre exemple : la relativité donne le moyen de rajeunir, ou, plutôt, de vieillir moins vite. Comment? Vous vous embarquez en 1966 dans un vaisseau cosmique (encore!) à la vitesse de 299 000 km/s, vous faites un voyage intersidéral pendant deux ans, vous revenez sur la Terre et vous trouvez les gens vieillis de... cinquante ans.

Encore un autre phénomène bizarre affirmé par ces messieurs les relativistes : plus un objet passe rapidement devant vous, plus il vous apparaît aplati, et, si un observateur est sur cet objet, il vous verra vous aussi contracté. N'y a-t-il pas là de quoi nous rendre inquiets et sceptiques?

En revanche, les Mémoires d'Einstein publiés en 1916 à propos de la relativité générale militeraient en faveur de la seconde hypothèse. Ils nous apprennent qu'il faut considérer notre univers comme courbe, que la ligne droite n'est pas le plus court chemin d'un point à un autre. A l'appui de tout cela, des calculs ardus, des entités mathématiques nouvelles : les tenseurs ...Brrr!

En gens de bon sens, ayant bien les pieds sur terre, tournons le dos à de telles billevesées. La mécanique classique de Newton n'a-t-elle pas du bon? N'est-ce pas grâce à elle que nos constructeurs font de solides et confortables automobiles, que les avions à hélices ou à réaction sillonnent les airs, que les cosmonautes eux-mêmes réalisent quelques tours de Terre à la vitesse de 36 000 km/h?

Le touriste, cependant, en descendant prudemment le col de la Faucille pour éviter les mauvais tours que peut jouer la force centrifuge, aperçoit près de Genève les installations imposantes du CERN, où fonctionne un synchrotron. Là, les physiciens étudient le comportement de particules extrêmement rapides, dont la vitesse avoisine celle de la lumière. Eh bien! la mécanique qui régit leur mouvement est obligatoirement relativiste et, sans le secours des formules établies par Einstein en 1905, les ingénieurs auraient été incapables de concevoir et de faire fonctionner cet appareil.

Ce même touriste, encore, sera perplexe en approchant des installations de Marcoule. Il sait par ouï-dire qu'on fabrique là de l'électricité à partir de l'uranium. Tout le secret de ces grandes usines réside dans une formule, $W = mc^2$, découverte en 1905, qui prouve que la matière est un réservoir énorme d'énergie. $W = mc^2$ est également une formule relativiste. Einstein, à la fin de sa vie, n'a-t-il pas été lui-même terrifié par certaines conséquences de sa théorie (bombes atomiques et thermonucléaires)?

La relativité, c'est sérieux. Bien que les disciples d'Einstein aient été rares au début, on peut dire qu'actuellement tous les physiciens, sans exception, sont résolument relativistes. Les formules fondamentales de la doctrine n'apparaissent-elles pas jusque dans les manuels de physique du second degré?

Cette nouvelle théorie est donc née avec le XXe siècle. Elle a tenté de réconcilier deux sciences qui avaient évolué chacune de son côté, en s'ignorant l'une l'autre : la mécanique et l'électromagnétisme. En quelques mots, voici le désaccord : mécanique et électromagnétisme n'admettaient pas les mêmes invariants, c'est-à-dire n'admettaient pas les mêmes formules de transformation quand on changeait de repère. Les savants, comme tous les hommes épris d'unité, trouvaient ce désaccord profondément choquant.

Einstein, pour les réconcilier, entreprit de critiquer de très près les notions fondamentales de temps et d'espace.

Jusqu'alors les scientifiques considéraient le temps et l'espace comme deux notions totalement distinctes : mettre sa montre à l'heure d'une part, repérer sa position sur un plan de Paris d'autre part, sont deux opérations « évidemment » indépendantes. Les physiciens, en outre, croyaient à un temps et un espace « absolus ».

Ces postulats classiques ont été rejetés par Einstein et ses disciples : l'espace et le temps ne sont ni absolus ni indépendants. Il faut en faire une seule entité, l'espace-temps, qui, seul, a un caractère invariant. Chaque observateur découpe cet espace-temps en espace et en temps à sa manière, et nous verrons que ce découpage est différent pour deux observateurs en mouvement l'un par rapport à l'autre. Cette notion (de relativité de l'espace et du temps) constitue l'essentiel de la doctrine que nous nous proposons de développer.

Certaines conséquences, comme celles qui sont énoncées au début, en paraîtront étranges. Mais, rassurons-nous; des vérifications expérimentales multiples, sérieuses et précises les ont confirmées.

Aussi allons-nous débuter par un examen critique des notions de temps et d'espace. Des contradictions apparaîtront, qui nous obligeront à reconstruire progressivement ces notions et nous feront aboutir finalement à cette belle synthèse qu'est la relativité générale.

espace et temps absolus : faut-il croire Newton ?

La notion de temps remonte aux origines de l'humanité. Les premiers hommes observèrent la suite des jours et des nuits, le retour périodique des phases de la Lune, l'alternance des saisons, les battements réguliers du pouls... Les phénomènes naturels se succèdent donc dans un ordre déterminé. Certains même se reproduisent identiques à eux-mêmes.

Mais, attention! Il se peut qu'il soit difficile de préciser l'ordre de succession de deux phénomènes. En effet, le critère le plus sûr pour affirmer qu'un événement est postérieur à un autre est de démontrer qu'il en est la conséquence (c'est-à-dire que le premier est la cause du second). Nous dirons, par exemple, que nous sommes allés chez le médecin après être tombés malade, puisque c'est le fait même d'être malade qui nous a incités à aller nous faire soigner : la maladie est la cause de la visite et donc antérieure à elle. Nous comprenons, en revanche, qu'il puisse être délicat de juger si la naissance d'une étoile dans la nébuleuse d'Andromède est postérieure ou antérieure à la naissance de Charlemagne, car aucun lien de causalité n'existe entre deux phénomènes aussi disparates. De plus, la nébuleuse d'Andromède et le berceau de l'empereur d'Occident sont à une distance énorme, du reste fort mal connue, l'une de l'autre. Et le seul courrier permettant d'informer la Terre de l'événement stellaire est la lumière, qui, ne se propageant qu'à la vitesse de 300 000 km/s, met plusieurs années pour remplir sa mission.

Très tôt, les hommes cherchèrent à mesurer le temps : les Assyriens se servaient de gnomons, les Romains utilisaient des clepsydres. Galilée, au XVIIe siècle, découvrit les lois du pendule, en observant, selon la légende, les mouvements oscillatoires d'un lustre de la cathédrale de Pise. Nous pensons qu'il se servit des battements de son pouls pour vérifier expérimentalement l'égalité de durée (isochronisme) des petites oscillations du lustre. De ces travaux naquirent les premières horloges à balancier. Nous n'avons pas l'intention de décrire les progrès de la chronométrie, mais nous voudrions bien préciser ce qu'est un appareil qui mesure le temps; c'est un système utilisant un phénomène périodique découpant le temps en intervalles, et l'on admet, par définition même, que ces intervalles sont égaux. (Il se peut, d'autre part, que des étalons de durée soient nécessaires. Par exemple, on pourrait utiliser la transmutation en proportion fixée d'une masse arbitraire d'émanation du radium.)

le temps absolu de Newton

Progressivement, la notion de temps est sortie de sa gangue intuitive pour se préciser et aboutir à la description mathématique des phénomènes. Newton, le premier, a essayé d'introduire une distinction entre le sens vulgaire de temps et son sens absolu, vrai, mathématique. Il écrivait dans ses célèbres *Principia :* « Le temps absolu, vrai et mathématique, sans relation avec rien d'extérieur, coule uniformément et s'appelle « durée ». Le temps apparent et vulgaire est cette mesure possible et externe d'une partie de durée quelconque, égale ou inégale, prix du mouvement : telles sont les mesures d'heures, de jours, de mois, etc., dont on se sert ordinairement, à la place de temps vrai », et, ailleurs : « Il est très possible qu'il n'y ait point de mouvement parfaitement égal qui puisse servir à la mesure exacte du temps ; car tous les mouvements peuvent être accélérés, mais le temps absolu doit toujours couler de la même manière. »

Ces phrases, qui ne sont d'ailleurs qu'une traduction du texte latin, peuvent paraître quelque peu confuses et hermétiques. L'idée profonde du grand savant est de montrer l'existence d'un temps absolu, valable pour l'univers entier. Newton l'érige en dogme, en concept dominant toutes les choses de ce bas monde. Faute de mieux, il le dit sans relation avec le monde extérieur. Comme il est malgré tout un homme de science, il sent la nécessité de mesurer ce temps idéal; la mesure du temps absolu, il l'appelle le « temps apparent, vulgaire », et il note bien que cette mesure ne pourra se faire qu'en observant des mouvements (par exemple, les oscillations d'un pendule, la rotation de la Terre). Et, tout de suite, un doute naît dans son esprit : et si tous les mouvements observables n'étaient pas uniformes, si aucun ne se reproduisait exactement à des intervalles de temps égaux, il serait impossible d'atteindre le temps absolu. Mais Newton balaie ce doute : même si le temps absolu n'est pas mesurable, il existe. Le grand savant croit au vieux dieu Cronos, imperturbable, mangeant consciencieusement ses enfants dès qu'ils naissent (et pourtant Zeus vivra!). En mécanique classique, il existe bien un seul temps ***t;*** tous les mathématiciens le désignent par la même lettre; les lycéens se débattent avec lui dès les classes du premier cycle. C'est le même, où que l'on soit, sur la Terre, sur la Lune, à bord d'un vaisseau spatial. En un mot, **c'est un invariant fondamental.**

Mieux armés que Newton, nous sommes à même de pousser plus loin les problèmes que pose la mesure du temps. Transportons-nous au cap Kennedy (ou à son équivalent soviétique) : un « rampant » et un cosmonaute possèdent chacun un chronomètre de précision; étant l'un à côté de l'autre, ils peuvent les mettre à la même heure, les synchroniser. Si le rampant, comme il se doit, reste au sol, tandis que son camarade part en fusée pour un voyage interplanétaire, le chronomètre de ce dernier sera constamment à la même heure que celui du Terrien, du moins si l'on en croit Newton.

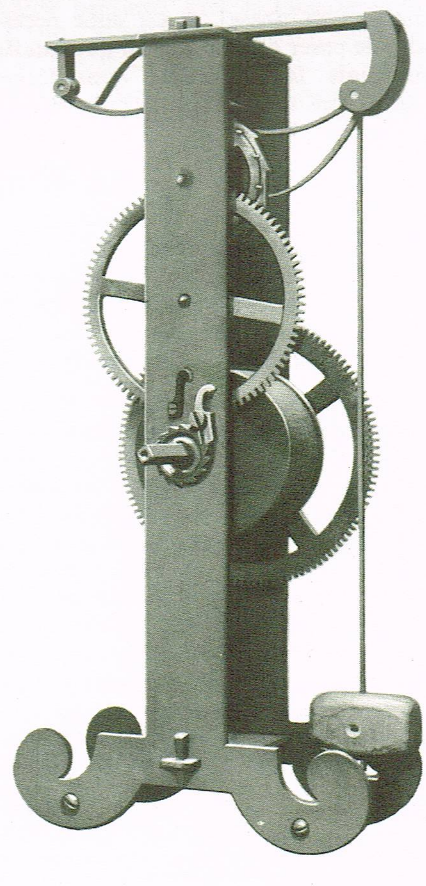
Comment nos deux amis pourront-ils s'assurer que leurs chronomètres sont à la même heure? Imaginons, par exemple, que le cosmonaute, au terme de son voyage, se pose sur notre satellite, la Lune, et y établisse une station d'observation. Même s'il croit, comme Newton, au temps absolu, il pourrait se faire qu'il ait des doutes sur la régularité de son chronomètre : il voudra la vérifier en communiquant avec son camarade resté sur la Terre.

Le Parisien, désireux de mettre sa montre à l'heure, appelle tout bonnement ODE 84-00. Notre futur Sélénite (habitant de la Lune) n'a pas d'autres ressources, s'il ne veut pas subir un voyage de retour pour le seul plaisir de vérifier son chronomètre, que de communiquer avec la Terre par ondes électromagnétiques ultra-courtes. C'est le seul courrier à sa disposition. Et nous savons que sa vitesse est très grande certes, mais non infinie. Il lui faudra donc tenir compte du temps de propagation des ondes. Pour une distance entre les deux astres de 370 000 km (nous supposons

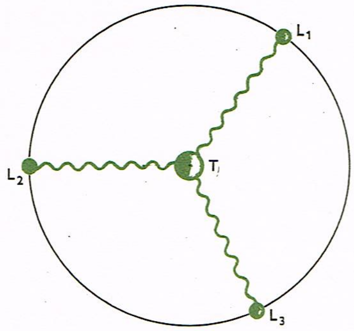
Gnomon de Tcheu-Kong. Le soleil passe par la fenêtre carrée, en haut de la tour, et projette une ombre qui glisse sur la longue murette horizontale. La longueur de l'ombre, variable avec l'heure et avec l'époque de l'année, est interprétée au moyen de tablettes.



Phot. Alinari-Giraudon.



Pendule construit d'après un dessin fait par le fils de Galilée sur les indications de son père.



Si la vitesse c de la lumière est constante, le temps mis par cette dernière pour parcourir TL_1, TL_2 et TL_3 sera constant (L_1, L_2 et L_3, différentes positions de la Lune dans son mouvement autour de la Terre, T).

pour simplifier que la Lune décrit un cercle ayant la Terre pour centre), ce temps s'élève à 2,5 secondes. Correction énorme, puisque l'on sait actuellement repérer les temps au dix-millième de seconde (en se servant d'un oscillographe cathodique) et les durées de trajet au dix-milliardième de seconde (mesures de phase). Il est donc nécessaire de poser comme certain que **quelle que soit la direction considérée, quelle que soit l'époque considérée, la vitesse des ondes électromagnétiques est constante. En particulier, elle ne doit pas dépendre de la vitesse de la Terre.** En revanche, s'il n'existe pas de courrier idéal (de vitesse constante), nous sommes dans l'impossibilité de tenir compte des temps de parcours, et nous ne pourrons donc pas régler des horloges à distance (1).

l'espace et sa conquête

Comme la notion de temps, la notion d'espace est délicate à dégager. Le bébé dans son berceau apprend à connaître l'espace : en agitant les bras, il se rend compte peu à peu à quelle distance se trouve son hochet; petit à petit, il fait connaissance avec les objets qui l'entourent, il meuble son voisinage proche. A deux ans, il pense pouvoir attraper la Lune, sans se douter de son éloignement considérable. Les hommes n'ont-ils pas, de même, cru longtemps que le Soleil, les planètes et les étoiles étaient tous à la même distance de la Terre? Le citadin novice faisant une course en montagne commet des erreurs analogues : la cime neigeuse qui brille au soleil lui paraît « à portée de sa main », et il faut des heures de marche pour l'atteindre. Seul, l'alpiniste averti, qui de nombreuses fois a gravi avec peine de longues pentes, sait apprécier la distance et évaluer les longues heures nécessaires pour vaincre le sommet convoité.

Ainsi, les hommes ont reconnu progressivement les positions et les distances des objets qui les environnaient. Intuitivement, ils ont acquis la notion d'espace. Les hommes se meuvent au milieu d'êtres, de choses qui les entourent; cet ensemble d'êtres, de choses constitue l'espace.

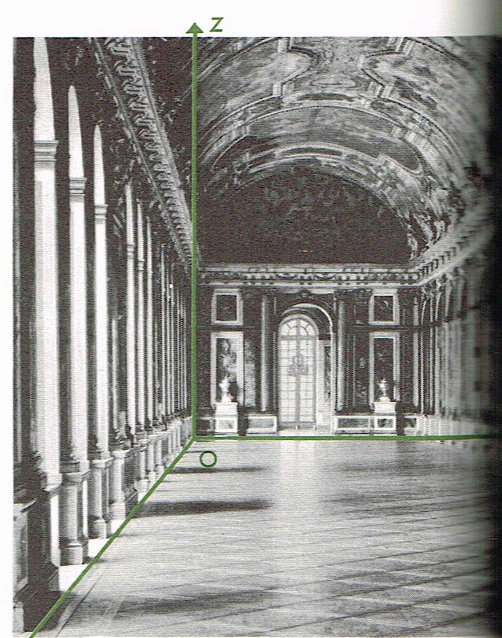
Puis un désir d'abstraction est venu; supprimer quelques arbres du paysage qui s'étend devant nous ne change pas fondamentalement le paysage. Et si l'on supprimait tout ce qui nous entoure, y aurait-il encore un espace? Qu'en pense un cosmonaute perdu au milieu des espaces interstellaires : il voit des étoiles; et on lui a appris combien elles étaient lointaines; il dira qu'entre lui et ces étoiles existe de l'espace vide. Voilà prononcé le grand mot : l'**espace vide!** L'espace est-il l'ensemble des êtres existants ou n'est-il que le cadre préétabli dans lequel prennent place ces êtres? Si nous suivons encore une fois Newton et ses disciples de la mécanique classique, il faut adopter le deuxième point de vue : l'espace en soi existe, il est absolu, il ne peut bouger; dans cet absolu, des corps se déplacent les uns par rapport aux autres et ce sont ces déplacements relatifs qui sont seuls mesurables. Newton énonce tout cela avec force, et il est bon de le citer encore une fois : « L'espace absolu, sans relation avec les choses externes, demeure toujours similaire et immobile. L'espace relatif est cette mesure ou dimension mobile de l'espace absolu, laquelle tombe sous nos sens par la relation avec les corps, et que le vulgaire confond avec l'espace immobile... » Il déclare ailleurs : « Tout est dans l'espace quant à l'ordre de la situation; c'est là ce qui détermine son essence et il serait absurde que les lieux absolus se mussent. »

Donc, pour Newton, l'espace absolu existe, et il est malheureux que les Grecs n'aient pas inventé le dieu de l'Espace : il aurait formé avec Cronos un couple très digne de la mythologie antique. La position d'un objet, le mouvement d'un objet doivent pouvoir être repérés par rapport à cet **espace absolu.** Newton, les physiciens du XVIII[e] et du XIX[e] siècle avaient la conviction profonde qu'aucun objet matériel, qu'aucun astre n'était fixe dans l'univers. Néanmoins, ils continuaient à croire à l'existence de l'espace absolu et à un moyen pour **repérer** les mouvements par rapport à un milieu, bien mystérieux d'ailleurs, « attaché » à l'espace absolu.

Qu'est-ce qu'un repère ?

C'est un solide de référence par rapport auquel on étudiera les positions et les mouvements des objets se déplaçant dans son voisinage. Dans une salle, par exemple, différents objets se meuvent, décrivent des trajectoires variées. Il est tout indiqué de choisir les murs de la pièce comme **repère,** comme **système de référence,** comme **référentiel,** ces trois expressions étant synonymes.

Evidemment, un tel référentiel n'est pas fixe, puisqu'il est solidaire de la Terre qui tourne autour du Soleil, qui lui-même est une étoile mobile de notre Galaxie, qui elle-même se déplace par rapport aux autres galaxies... Malgré cela, les physiciens d'avant 1900 espé-

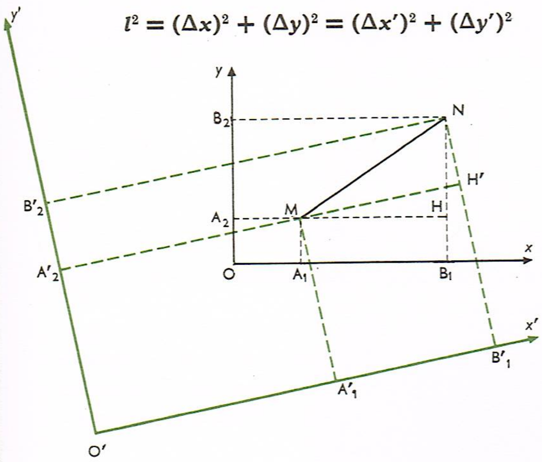


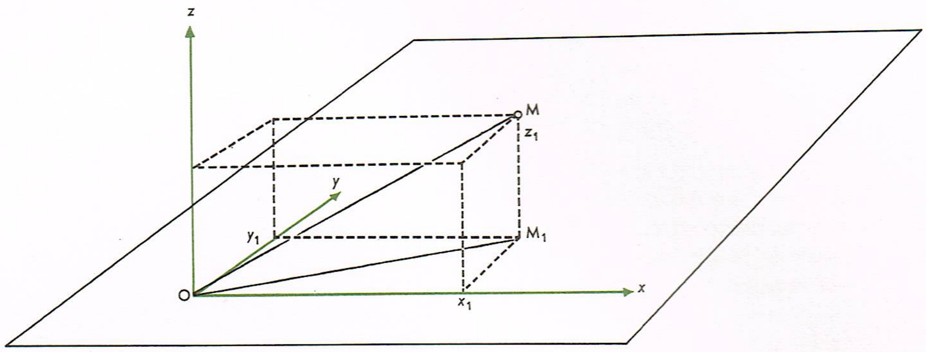
Les trajectoires suivies par tout ce qui se meu l'intérieur d'un local peuvent être étudiées en rapportant au référentiel constitué par les mu

Phot. Société des Amis de Versailles.

---

(1) D'ailleurs, il ne faut pas croire qu'il soit plus facile de régler l'une sur l'autre deux horloges terrestres, l'une, par exemple, placée à New York et l'autre à Paris. Les ondes électromagnétiques suivent un chemin zigzagant et variable. En outre, la propagation ayant lieu dans l'air, dont la composition et la densité électronique varient avec l'altitude, la vitesse des ondes n'est pas constante et le problème devient terriblement compliqué.





$OM^2 = OM_1^2 + M_1M^2 = (x_1^2 + y_1^2) + z_1^2$

raient toujours découvrir le « référentiel absolu ». La mécanique ayant échoué dans cette recherche, ils se sont tournés au XIX[e] siècle vers l'optique et l'électromagnétisme, mais n'anticipons pas. Avant de poursuivre, il faut bien préciser la notion fondamentale de référentiel. Sans cela, on ne peut saisir sérieusement le sens profond des théories relativistes.

Repérer la position d'un point M situé dans un plan est chose facile : il suffit de choisir un point O origine et de tracer deux droites rectangulaires Ox, Oy ; de M on abaisse les perpendiculaires MA_1 et MA_2 sur $x'Ox$ et $y'Oy$. La connaissance des distances $OA_1 = x_1$ et $OA_2 = y_1$ permet de repérer le point M par rapport au référentiel Oxy; x_1 et y_1 seront les **coordonnées** du point M par rapport à ce référentiel. Les deux nombres x_1 et y_1 suffisant à déterminer la position du point M par rapport au repère, nous dirons que le **plan est une multiplicité à deux dimensions.**

Dans un plan, une règle rigide de longueur l a pour extrémités M (de coordonnées x_1 et y_1) et N (de coordonnées, par rapport au même repère, x_2 et y_2). Le théorème de Pythagore permet d'écrire, pour le triangle rectangle MNH,

$$MN^2 = MH^2 + HN^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2,$$

ou, en employant les notations Δ pour les différences,

$$l^2 = (\Delta x)^2 + (\Delta y)^2.$$

Changeons maintenant de repère. Prenons un autre point origine O′ et deux autres axes rectangulaires O′x' et O′y'. Dans ce nouveau repère, M aura pour coordonnées x'_1 et y'_1, et N, x'_2 et y'_2, et l'on pourra écrire, cette fois-ci pour le triangle rectangle MNH′,

$$MN^2 = MH'^2 + H'N^2 = (x'_2 - x'_1)^2 + (y'_2 - y'_1)^2 = (\Delta x')^2 + (\Delta y')^2.$$

Nous avons donc

$$\mathbf{l^2 = (\Delta x)^2 + (\Delta y)^2 = (\Delta x')^2 + (\Delta y')^2.}$$

Que signifie cette formule ? Le carré de la longueur l de la règle MN s'exprime dans notre exemple de deux manières différentes suivant le référentiel adopté. Plus généralement, et de manière analogue, quel que soit le référentiel cartésien adopté (référentiel schématisé par deux axes rectangulaires Ox et Oy), la quantité

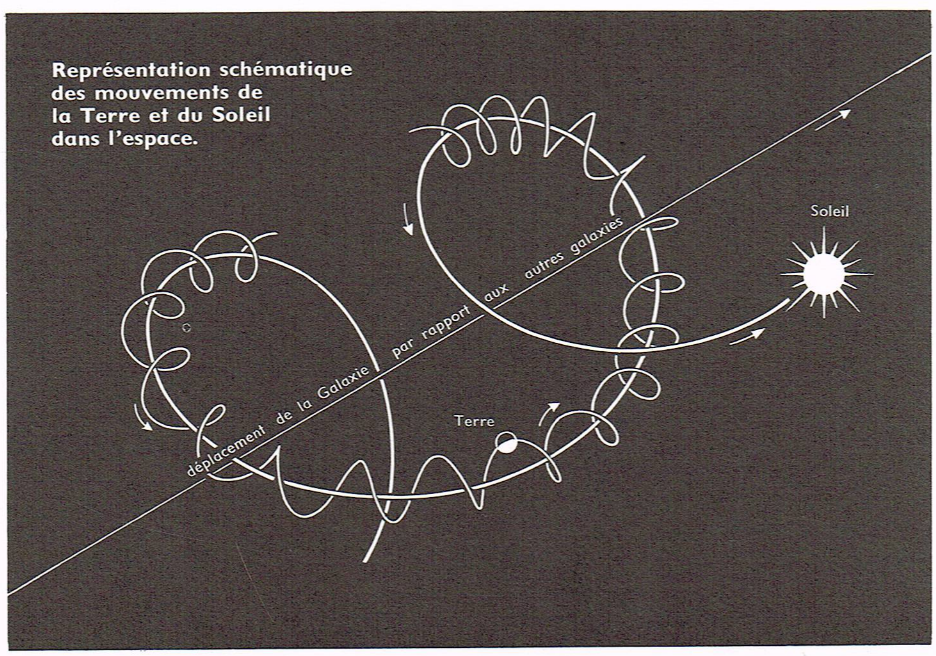
$$\Delta x^2 + \Delta y^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$$

reste la même, reste invariante, puisqu'elle représente la longueur de la règle. **Les valeurs x_1, x_2, y_1, y_2 varient quand on change de référentiel cartésien, mais $(x_2 - x_1)^2 + (y_2 - y_1)^2$ reste une grandeur constante.** La racine carrée de cet invariant, elle aussi invariante, s'appelle la distance de M à N (ou de N à M) :

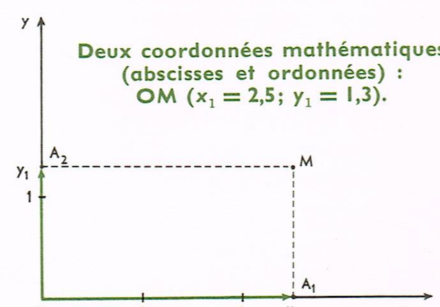
$$d\,(M, N) = d\,(N, M) = \sqrt{\Delta x^2 + \Delta y^2}.$$

Il est bien évident que deux coordonnées sont insuffisantes pour repérer la position d'un point dans l'espace. Il faut introduire une troisième dimension : **notre espace est une multiplicité tridimensionnelle.**

Prenons dans l'espace un plan de référence quelconque P et dans ce plan



Représentation schématique des mouvements de la Terre et du Soleil dans l'espace.



Deux coordonnées mathématiques (abscisses et ordonnées) : OM ($x_1 = 2{,}5$; $y_1 = 1{,}3$).

deux axes rectangulaires Ox et Oy. Projetons M en M_1 sur le plan P. M_1 a pour coordonnées x_1 et y_1. La connaissance de $MM_1 = z_1$ repère le point M par rapport au trièdre trirectangle $Oxyz$.

On montrerait très simplement que si les coordonnées des deux points M et N sont respectivement x_1, y_1, z_1, et x_2, y_2, z_2, le carré de la distance MN sera donné par la relation

$$MN^2 = l^2$$
$$= (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2,$$

c'est-à-dire

$$\boldsymbol{l^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2},$$

et qu'un changement de référentiel rectangulaire laisserait invariante la quantité

$$l^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$$
$$= (\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2,$$

ou sa racine carrée, appelée encore « distance » (dans l'espace à trois distances) des points M et N,

$$d\,(M, N) = d\,(N, M)$$
$$= \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}.$$

La distance ainsi définie de deux points de notre espace est un invariant fondamental par rapport à tous les systèmes de référence rectangulaires que nous pourrions choisir.

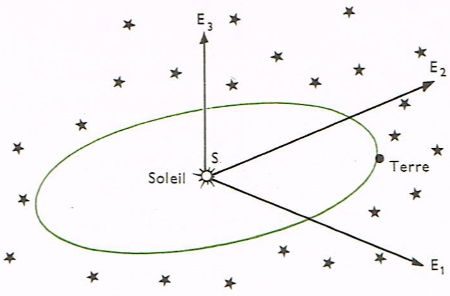
En particulier, la distance à l'origine du point M, de coordonnées x, y, z, vaut $OM^2 = l^2 = x^2 + y^2 + z^2$.

On peut encore dire que le vecteur OM, dont les trois composantes sont x, y, z, a pour longueur $\sqrt{x^2 + y^2 + z^2}$. Si l'on change de repère cartésien, tout en conservant la même origine O, les composantes du vecteur OM sont différentes, mais la quantité $x^2 + y^2 + z^2$ reste la même. Ainsi les vecteurs de l'espace à trois dimensions possèdent trois composantes : ce sont des **trivecteurs.** Si, par hasard, il nous venait à l'idée d'imaginer un « espace à quatre dimensions », les vecteurs de ce nouvel espace seraient caractérisés par quatre composantes, w, x, y, z, et ils seraient qualifiés de **quadrivecteurs.**

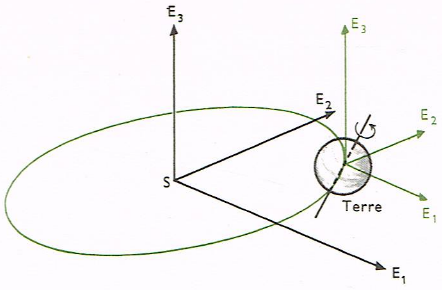
Les référentiels que nous venons d'envisager sont-ils immobiles dans l'espace absolu ? Si oui, nous pourrions alors les qualifier eux aussi d'« absolus ».

Mais est-il possible de mettre en évidence de tels référentiels ? Afin de tenter de répondre à cette question, il faut définir maintenant les référentiels galiléens.

Cherchons maintenant à adopter un référentiel qui soit commode pour repérer le mouvement des corps de toute



Le référentiel de Copernic.



Un référentiel de Galilée.

nature : projectiles, satellites lancés de la surface de la Terre, planètes tournant autour du Soleil, etc. Prendre le centre de la Terre comme origine conviendra mal pour étudier les mouvements des planètes autour du Soleil. A la suite de Copernic, il est normal de choisir précisément le centre du Soleil (ou, plus exactement, le centre de gravité du système solaire). Croyant — à la suite de Newton et des mécaniciens classiques — à la réalité de l'espace absolu, nous choisirons des axes OX, OY, OZ dont les directions sont fixes dans l'espace.

Nous nous heurterons dès ce moment à une grave difficulté : comment matérialiserons-nous ces directions fixes, puisque les astronomes ont prouvé que toutes les étoiles étaient mobiles les unes par rapport aux autres ?

Faute de mieux, nous dirons que les directions fixes choisies sont les directions de trois étoiles tellement éloignées de la Terre qu'on peut les considérer comme fixes en première approximation.

Un tel référentiel a été appelé **référentiel de Copernic** ; nous avons postulé que les directions SE_1, SE_2, SE_3 étaient « fixes », mais rien ne nous permet d'affirmer que le Soleil est fixe dans l'espace. Le référentiel de Copernic n'est sûrement pas le référentiel que nous recherchons.

Tout référentiel dont les axes restent parallèles à ceux de Copernic et dont l'origine se meut par rapport à S d'un mouvement rectiligne et uniforme est dit **référentiel de Galilée.** Par exemple, choisissons le centre de la Terre comme origine. Pendant un court instant nous pouvons considérer son mouvement par





Les sens de la vue
et du toucher
nous introduisent dans un
monde de mouvements.
Même bébé perçoit
l'emplacement des objets,
dans le sens propre,
et non des fonctions
dans le temps.
Espace et temps absolus
sont beaucoup moins
naturels et sensibles
que l'espace temps.
Les artistes ont bien compris
cette évidence.

●

A gauche, « Combat de coqs »
par Pignon (1959),
Galerie de France, Paris ;
à droite, « Pluie »
par Tore Johnson, Stockholm.

rapport au centre du Soleil comme rectiligne et uniforme. Nous aurons là un référentiel de Galilée lié au centre de la Terre; remarquons bien que la Terre tourne par rapport à ce référentiel.

Si l'on étudie les mouvements des projectiles à la surface de la Terre, on choisira un référentiel lié à la Terre, et la mécanique classique montre que des forces supplémentaires (forces de Coriolis) doivent être introduites pour expliquer complètement les phénomènes observés.

Evidemment, un référentiel galiléen n'est pas lui non plus absolu. Le référentiel de Copernic, les référentiels de Galilée ne sont pas fixes dans l'espace. Leur position varie avec le temps. **Dès maintenant il apparaît que les notions d'espace et de temps sont toujours intimement liées.** Si le référentiel absolu existait, il en serait tout autrement, car, ce référentiel étant fixe, il serait indépendant de la notion de temps.

la très classique transformation de Galilée

Tout référentiel G′ se déplaçant d'un mouvement rectiligne et uniforme par rapport à un référentiel de Galilée G sera donc lui aussi référentiel de Galilée. Il est utile de prendre un exemple simple, bien qu'approximatif.

Admirons le paysage ci-dessous. Les collines, le village, le sol de la route sont liés à la Terre. En négligeant en première approximation la rotation de notre planète, la route, les collines et le village définiront un référentiel G, que schématise le trièdre O*xyz*.

A midi, une auto passe au carrefour O, animée de la vitesse constante v, et se dirige sur la route rectiligne dans la direction Ox. L'auto est un solide (roues mises à part) animé d'un mouvement rectiligne et uniforme par rapport au référentiel galiléen O*xyz*. Elle permet de définir un trièdre O′$x'y'z'$ (évidemment O′x' est parallèle à la route, et de même O′y' parallèle à Oy, O′z' parallèle à Oz). Le référentiel « auto » G′ est bien un référentiel galiléen se déplaçant par rapport au référentiel galiléen « paysage » G.

Une compagnie pétrolière a installé un poste d'essence M au bord de la route. Sa distance au carrefour O est x = OM, tandis que sa distance à la voiture (point O′) est x' = O′M. Quelle relation y a-t-il entre x' et x?

Un chronomètre fixe est installé au point O et est déclenché à midi; il va donc mesurer le temps en prenant midi comme origine (midi est pris comme temps zéro). L'automobiliste possède lui aussi un chronomètre, et il le déclenche également à midi, moment où il franchit le carrefour O. Quand le conducteur lit le temps t à son chronomètre, il a parcouru la distance OO′ = vt. La relation cherchée est donc :

$$x' = x - vt.$$

Evidemment, les chronomètres, l'un resté au point fixe O, l'autre emporté par la voiture, indiquent le même temps :

$$t' = t.$$

Ces deux formules constituent ce que l'on appelle la « transformation de Galilée (1) »; on peut ainsi calculer l'abscisse du poste à essence M par rapport au référentiel galiléen « auto » G′, connaissant l'abscisse de M par rapport au carrefour O, lié au référentiel « paysage » G.

(1) Tout se passant sur l'axe Ox, nous n'avons pas eu à calculer les deux autres coordonnées y et z, qui restent toujours nulles. Plus généralement, il faudra ajouter deux autres relations entre (x, y, z) et (x', y', z').

premier contact avec la notion d'événement

Dès maintenant, les notions d'espace et de temps apparaissent intimement liées; la formule précédente le prouve nettement. Pour bien préciser notre naissance, il ne suffit pas de dire où nous sommes nés, mais aussi à quelle date nous sommes nés.

Autre exemple : le 14 juillet 1789, les Parisiens ont assiégé la Bastille. Par définition, le lieu du fait (Paris) et sa date (14 juillet 1789) constitueront un événement.

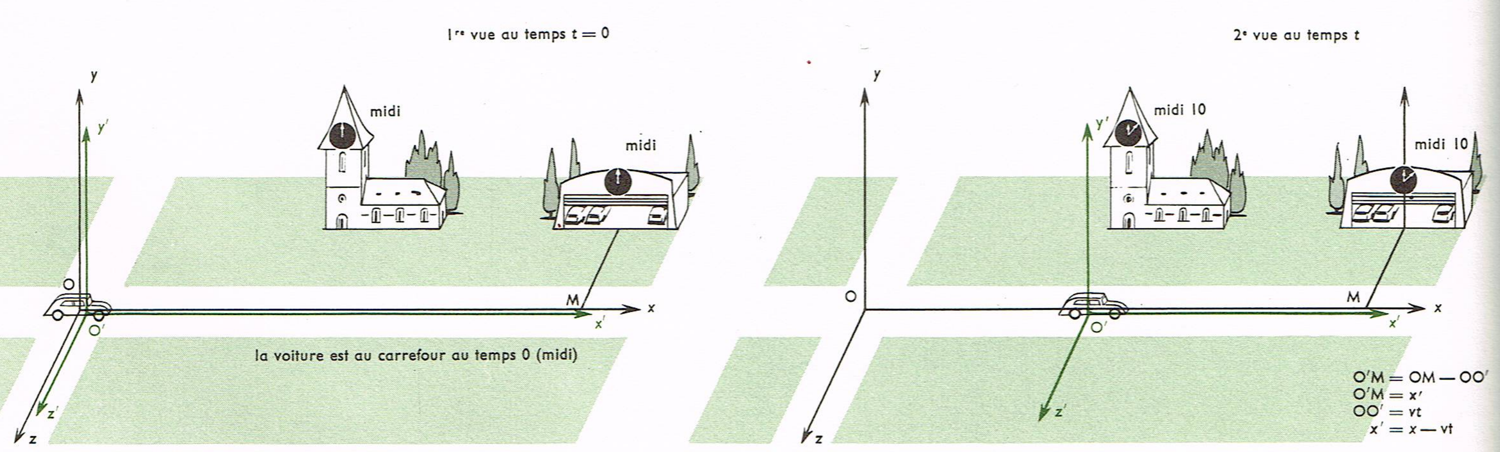
Un événement E est ainsi caractérisé par quatre nombres : x, y, z coordonnées du lieu par rapport à un référentiel, t temps compté à partir d'une certaine origine, généralement la naissance de Jésus-Christ. (En fait, les calculs faits pour la déterminer étaient faux et le Christ est né six ou sept ans avant le début de notre ère. Mais peu importe ici). Un second événement E′ se caractérise par quatre autres nombres x', y', z', t'. Evidemment, si l'on désire comparer deux événements, il est nécessaire de les rapporter à un même référentiel et de prendre une origine des temps commune. Par exemple :

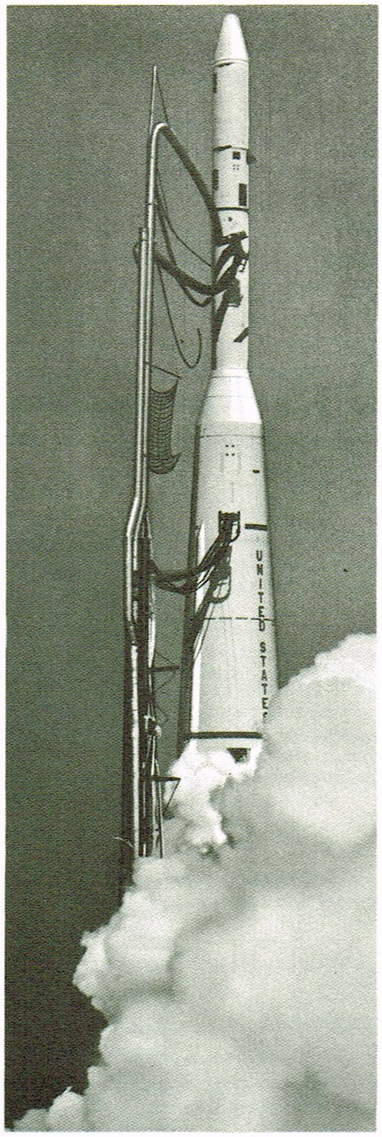
prise de la Bastille, 1789;

lancement d'un satellite au cap Kennedy, 1960.

A un tel couple sont associés deux invariants fondamentaux : leur distance dans l'espace (5 000 km) et l'intervalle de temps qui les sépare (171 années). Si l'on change de repère, la distance et l'intervalle de temps qui séparent les deux événement seront inchangés (2).

(2) Nous reviendrons sur cette notion d'événement en relativité, quand nous introduirons le concept d'espace-temps. En relativité, la distance et l'intervalle de temps séparant deux événements perdront leur caractère d'invariance.





la déficience congénitale de la mécanique

Un passager monte à Orly à bord d'un Boeing 707 assurant le service de nuit Paris-New York. Les réacteurs sont mis en marche, l'avion gagne le point fixe en bout de piste. A ce moment, et malgré le bruit assourdissant, l'homme s'endort... Il se réveille deux heures après, l'avion est au-dessus de l'Atlantique, dans la nuit profonde, à 10 000 m d'altitude, et vole d'un mouvement parfaitement rectiligne et uniforme. Si le voyageur en question n'a pas conscience d'avoir dormi, il lui est impossible de savoir s'il est toujours arrêté sur la piste d'Orly ou s'il est en mouvement rectiligne et uniforme au-dessus de l'Océan. L'avion est un référentiel galiléen S′, qui se déplace par rapport au référentiel « Terre » galiléen (ou du moins, comme nous l'avons vu précédemment, approximativement galiléen). Si le passager est quelque peu physicien, les expériences mécaniques qu'il pourra tenter à l'intérieur de l'avion conduiront à des résultats rigoureusement identiques à ceux qu'il obtiendrait à terre : les corps tomberont dans l'avion suivant la verticale, animés de la même accélération, il retrouvera les mêmes lois du pendule simple, etc.

Que conclure de cet exemple ? Les lois de la mécanique sont les mêmes quel que soit le repère galiléen choisi pour les étudier. **Il est impossible par des expériences mécaniques faites à l'intérieur d'un système clos de mettre en évidence son mouvement de translation uniforme ;** la notion de translation uniforme n'a pas de caractère absolu. Donc, les phénomènes mécaniques ne permettront pas d'atteindre ce repère absolu que nous poursuivons sans cesse. L'espace absolu échappe aux investigations expérimentales de la mécanique.

●

Pour mieux saisir encore cette équivalence mécanique des divers référentiels galiléens, allons rendre visite à un porte-avions en manœuvre.

Le navire vogue sur une mer d'huile, les pilotes qui vont décoller et atterrir sur la piste d'envol sont encore novices ; aussi le commandant de bord a-t-il transmis l'ordre aux officiers mécaniciens de veiller à conserver au bâtiment une vitesse bien constante. Ce bateau est un référentiel galiléen S′ en translation uniforme par rapport au système galiléen « Terre » S (dans ces exemples



171 années d'écart et 5 000 kilomètres...

Phot. U. S. I. S.

simples, la rotation de notre planète est toujours négligée).

Le porte-avions vogue à une vitesse de 32 nœuds, soit environ 60 km/h. Un avion quitte la piste d'envol par l'avant, il accélère progressivement pour atteindre la vitesse $u' = 200$ km/h par rapport au navire. Sa vitesse u par rapport à l'Océan est donc de $200 + 60 = 260$ km/h. u et u' sont différentes, et $u = u' + v$. Mais que dire des accélérations γ et γ' de l'avion par rapport au bâtiment et par rapport à la mer? Elles sont égales! En effet, l'accélération est l'accroissement de vitesse pendant une seconde; d'une part, les vitesses u et u' de l'avion sont distinctes, mais leur variation est la même puisque leur différence v est une quantité rigoureusement constante (1).

D'autre part, il n'y a aucun doute (nous nous repentirons dans quelques instants de cette affirmation péremptoire) que les chronomètres embarqués fournissent les mêmes indications que ceux des stations terrestres : **la seconde « porte-avions » et la seconde « Terre » sont identiques.** Quelle est la cause de l'accélération de l'avion? C'est la force $\vec{F}$ de poussée du réacteur, tout à fait indépendante du mouvement du navire. En admettant que la masse m de l'avion soit une grandeur indissolublement associée à l'appareil, le principe fondamental de la dynamique $\vec{F} = m\vec{\gamma}$ s'exprime exactement de la même manière par rapport au porte-avions ou par rapport à la Terre. **La formule fondamentale**

(1) $\gamma' = \frac{du'}{dt} = \frac{d}{dt}(u + v) = \frac{d}{dt}u = \gamma$

(puisque v est constant, sa dérivée est nulle. En outre, u, u' et v étant parallèles, on peut opérer sur les vecteurs ou sur leurs mesures).

la vitesse de la lumière

L'étude des ondes électromagnétiques fait apparaître une constante fondamentale c, dont les dimensions sont celles d'une vitesse. Cette grandeur est identifiable à la vitesse de propagation dans le vide — on dit encore la « célérité » — d'une perturbation électromagnétique.

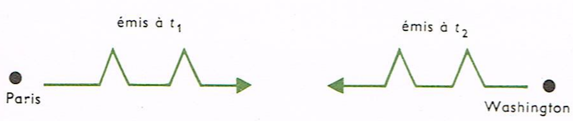
Les mesures effectuées sur les ondes hertziennes et sur les ondes lumineuses, au moyen de méthodes électriques ou optiques tout à fait indépendantes, multipliées depuis le XVIIe siècle, ont conduit à une valeur de c voisine de 300 000 km/s, la valeur la plus probable étant déterminée à 1 km/s près environ. Aucune mesure ne permet, actuellement, de déceler une quelconque variation de c, en fonction du temps par exemple.

Les principes de quelques-unes des méthodes mises en œuvre sont ici représentés schématiquement.

Les deux groupes de techniques correspondent à des régions différentes du domaine des fréquences électromagnétiques.

vitesse de propagation des ondes hertziennes

Mesure directe (Abraham et Ferrié, 1912).



Des séries de signaux brefs (impulsions) sont émises de Paris et de Washington. Le signal émis à Paris au temps t_1 est reçu à Washington au temps t'_1.

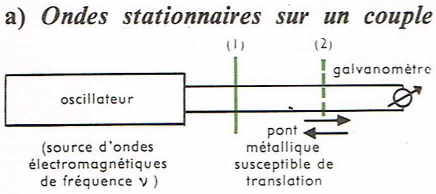
Le signal émis à l'instant t_2 à Washington est reçu à Paris à l'instant t'_2.

La durée moyenne du trajet est donc $\frac{(t'_1 - t_1) + (t'_2 - t_2)}{2}$.

Compte tenu de la distance Paris-Washington, la détermination de la vitesse moyenne de propagation, déduite d'un grand nombre de mesures portant sur la durée du trajet, donne $c \sim 296\,000$ km/s.

Mesures indirectes

a) *Ondes stationnaires sur un couple de fils parallèles (Hertz, Blondot, Mercier, 1923, etc.).*

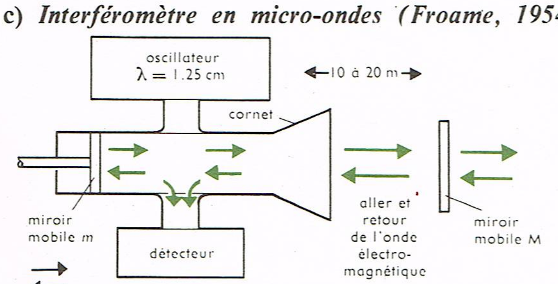


Il s'établit un système d'ondes stationnaires. Les maximums d'amplitude sont détectables au moyen du « pont » et d'un appareil de mesure. Deux maximums (1) et (2) sont séparés par la distance $\frac{\lambda}{2}$. La vitesse de transmission, tirée de la relation $\lambda\nu = c$, est donnée à une dizaine de km/s près.

b) *Mesures de la capacité d'un condensateur en unités électromagnétiques et électrostatiques (Dorsey, 1906)*

Les relations établies dans le chapitre sur les équations de Mawxell fournissent $c = \sqrt{\mu_0 \varepsilon_0}$. Si la perméabilité μ_0 est égale à l'unité, la détermination de la constante diélectrique ε_0 entraîne immédiatement celle de c.

c) *Interféromètre en micro-ondes (Froame, 1954).*



Lorsque le miroir M est déplacé suivant le faisceau issu du cornet, le détecteur rend compte des maximums résultant de sa recombinaison avec le faisceau réfléchi sur m. Ces maximums sont, comme sur les fils, distants de $\frac{\lambda}{2}$. L'incertitude sur la mesure de c (déduite de $c = \lambda\nu$) est inférieure à 1 km/s.

de la dynamique $\vec{F} = m\vec{\gamma}$ est invariante par rapport à tout changement de référentiel galiléen.

L'exemple du passager du Boeing nous conduit intuitivement à l'équivalence mécanique des systèmes galiléens, celui du porte-avions nous a permis de démontrer cette équivalence. Il est à souligner que cette démonstration repose sur des postulats qu'il faut bien préciser :

a) **Le temps mesuré est le temps absolu, identique pour tous les référentiels galiléens ;**

b) **La loi de composition des vitesses est une loi additive : $\vec{u'} = \vec{u} + \vec{v}$;**

c) **La masse d'un point matériel est un invariant qui lui est associé.**

Faisons le point : où en sommes-nous ? Les repères, avions, bateaux, fusées, ont défilé devant nos yeux. La tête nous tourne...

A quoi, avec Newton et les mécaniciens classiques, croyons-nous ?

a) **Au temps et à l'espace absolus ;**

b) **A l'équivalence mécanique des référentiels galiléens.**

Belle avance, puisque ces systèmes ne permettent pas de mettre en évidence le mouvement absolu !

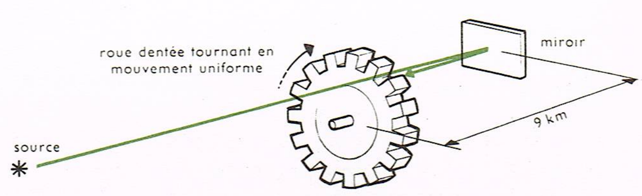
Il faut être réaliste : si le mouvement absolu existe, si ce concept a un sens, des expériences doivent le mettre en évidence. La mécanique classique se déclare incompétente. Force est bien de se tourner vers les phénomènes lumineux et électromagnétiques. Tel a été le cheminement de la pensée des physiciens du XIX[e] siècle, mais les expériences d'optique de Michelson ont définitivement ruiné leurs conceptions vers 1900.

vitesse de propagation des ondes lumineuses

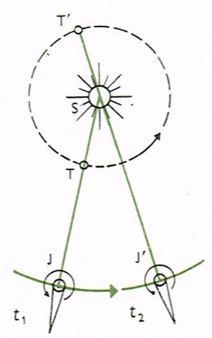
Au début du XVII[e] siècle, Galilée pressentait que la transmission n'était pas instantanée, mais il avait espéré que deux observateurs, placés à quelques kilomètres l'un de l'autre et repérant le moment où ils occulteraient leurs lanternes, pourraient donner approximativement la vitesse de la lumière. Mais les deux observateurs étaient-ils capables d'apprécier le cent-millième de seconde?

Deux siècles et demi plus tard, Fizeau « améliore » les conditions dans son expérience célèbre.

Les impulsions de la lumière qui, à l'aller, passent entre deux dents de la roue, ne seront pas occultées au retour, après réflexion sur le miroir, si l'on se retrouve dans une situation analogue, c'est-à-dire dans un nouveau creux. Le temps écoulé pendant le trajet roue-miroir-roue est donc mesurable (la vitesse de la roue dentée et le nombre de dents étant fixés).



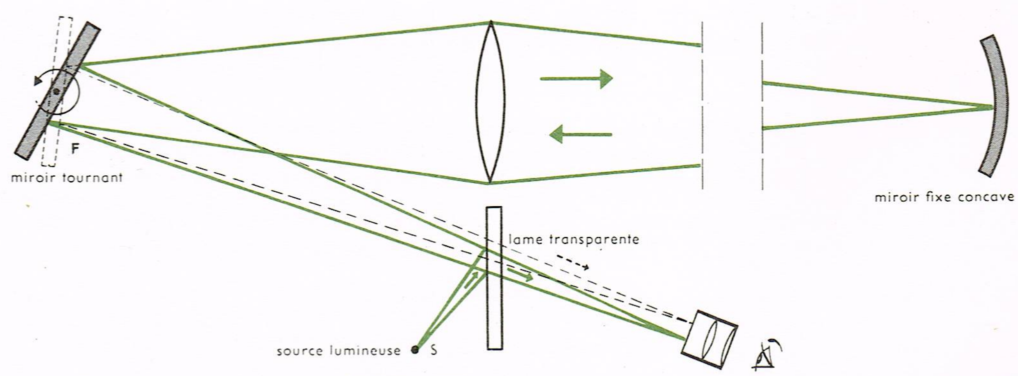
la méthode astronomique de Rœmer



Discutant des tables de calcul pour l'observation des satellites de la planète Jupiter, Rœmer, alors à l'Observatoire de Paris (1675), découvrit que la durée apparente de révolution des satellites en question, la date de leurs éclipses variaient suivant l'éloignement de Jupiter. Au temps t_1, la Terre est en T et Jupiter en J, tandis qu'au temps t_2 la Terre est en T', plus éloignée de Jupiter, alors en J'. Le trajet parcouru par la lumière venant du satellite est donc plus grand. Les écarts entre les temps d'immersion des satellites dans le cône d'ombre donnent ainsi la vitesse c, de l'ordre de 310 000 km/s d'après Rœmer.

le miroir tournant de Foucault (repris par Michelson en 1926)

Le faisceau de lumière issu de la source s'est réfléchi une première fois sur la face F du miroir tournant, puis sur le miroir fixe. Lorsqu'il est de retour sur F, celle-ci a tourné d'un certain angle et l'image de S est décalée par rapport à S. Une relation simple peut être mise en évidence entre ce décalage, la vitesse de rotation du miroir tournant (de l'ordre de 400 à 500 tours à la seconde pour les dispositifs actuellement utilisés dans les manipulations à l'usage des étudiants) et la vitesse c de la lumière. Celle-ci est déterminée à quelques kilomètres à la seconde près.



à la recherche du référentiel absolu

Nous voici donc à la recherche du repère absolu, fixe dans l'univers.

Des expériences purement mécaniques ne permettant pas de l'atteindre, comme nous venons de le voir, essayons de nous adresser à des **phénomènes électromagnétiques,** dont la lumière est une manifestation. La lumière se propage dans les milieux matériels transparents et dans le vide; dans le vide, sa vitesse est maximale et atteint 300 000 km/s.

l'éther et son mystère : l'expérience de Michelson

La lumière étant de nature ondulatoire, il était logique d'admettre un milieu élastique vibrant, permettant la propagation des ondulations lumineuses, tout comme les ondes sonores se propagent dans l'air, milieu élastique vibrant, tout comme les ondes sismiques se propagent dans le milieu élastique vibrant formé par la Terre. Avec beaucoup de prudence, même de méfiance, les physiciens du XIXe siècle imaginèrent un tel milieu. Il devait avoir des propriétés bizarres : extrême rigidité, puisque la vitesse de propagation des ondes lumineuses est très grande, extrême fluidité, pour ne pas opposer de résistance appréciable aux corps en mouvement. Il devait baigner tout l'espace, vide ou rempli de matière : les physiciens ont appelé ce milieu **éther.** Toute cette théorie, péniblement établie, fit faillite rapidement, et on ne trouve aucune interprétation mécanique raisonnable des propriétés du mystérieux véhicule des ondes lumineuses. Petit à petit, l'idée d'identifier l'éther à l'espace lui-même se fit jour : en tout point de l'espace peut exister un champ électromagnétique, et les équations de Maxwell prouvent que ce champ se propage à la vitesse $c = 300\,000$ km/s.

Au XIXe siècle, les physiciens disaient que des expériences optiques ou électromagnétiques devaient permettre de révéler un mouvement de translation par rapport à l'éther. Ainsi, l'éther, solide hypothétique, pourrait jouer le rôle de repère absolu, jusqu'ici insaisissable. Nous énoncerons la même proposition dans un langage différent : il se peut que de telles expériences mettent en évidence un mouvement de translation par rapport à l'espace lui-même, c'est-à-dire un mouvement de translation absolu.

Une équipe de physiciens se trouve dans un navire intersidéral, loin de toute étoile. A un certain instant, les réacteurs sont stoppés. Aucune force n'agissant sur le véhicule, ce dernier continue sa course animé d'un mouvement rectiligne et uniforme à la vitesse v.

Aucune expérience de mécanique ne permet de mettre en évidence ce mouvement. Mais des expériences d'optique ne pourraient-elles être positives?

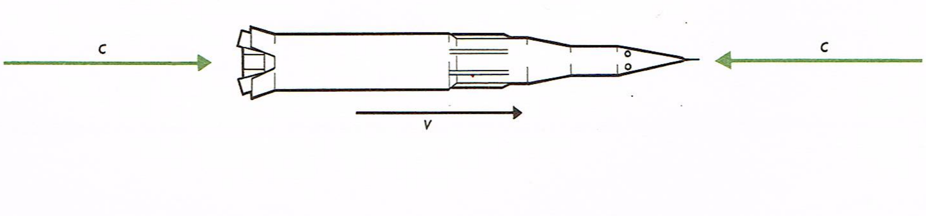
Un rayon lumineux vient au-devant du vaisseau spatial. Son milieu de propagation est l'éther ou l'espace absolu lui-même, sa vitesse est égale à c. Si nos physiciens, dans leur vaisseau, mesurent la vitesse de la lumière par rapport à eux, ils doivent trouver $c + v$. En revanche, ils doivent trouver $c - v$ pour tout rayon lumineux arrivant par l'arrière du véhicule. Ainsi, tout physicien en mouvement doit pouvoir constater que la vitesse de la lumière varie suivant sa direction de propagation : il pourra donc mettre en évidence son mouvement absolu; d'une manière imagée, on pourra dire qu'il a senti le vent d'éther.

Est-il besoin d'imaginer un laboratoire spatial pour réaliser cette expérience? Notre bonne vieille Terre suffit : elle se déplace dans l'espace à la vitesse de 30 km/s sur son orbite; pendant la durée d'une expérience, on peut admettre que son mouvement est de translation uniforme. Effectivement, l'expérience a été réalisée en 1890 par le physicien américain Michelson, qui désirait mettre en évidence le mouvement absolu de la Terre dans l'espace.

Le navire intersidéral précédent n'est donc autre que la Terre. Une question vient alors tout de suite à l'esprit : la source lumineuse utilisée sera-t-elle installée sur le vaisseau même (la Terre) ou bien faudra-t-il prendre un rayon lumineux issu d'une étoile lointaine?

Pour répondre à cette question, il faut savoir si la célérité de la lumière dépend de la « vitesse » de la source émettrice, évidemment mesurée par rapport à l'espace absolu. Les relativistes précisent que quelle que soit la vitesse de la source émettrice la célérité de la lumière est constante. Cette affirmation a été âprement discutée par les antirelativistes vers 1910. Il est donc nécessaire d'apporter une preuve expérimentale à l'appui de l'affirmation précédente; nous la trouverons dans les observations du mouvement des étoiles doubles : ces étoiles tournent l'une autour de l'autre suivant les lois de la gravitation universelle et les astronomes ont vérifié que leurs observations étaient en accord complet avec la loi de Newton. Cette vérification impose la constance de la vitesse de la lumière émise par les deux étoiles,

Pour l'équipage d'un astronef en vol balistique, aucune expérience mécanique ne saurait mettre en évidence le mouvement de l'engin. Seule, la vitesse relative de la lumière aurait pu permettre de constater ce mouvement à bord, puisque, le rayon c venant de l'avant, on devrait trouver une vitesse résultante égale à $c + v$, tandis que pour le rayon venant de l'arrière la vitesse serait $c - v$. Il n'en est rien, comme on a pu le constater depuis Michelson.



puisque l'information transmise par la lumière arrive sur Terre avec un retard considérable certes, mais rigoureusement constant. Donc, l'origine de la lumière importe peu : la source lumineuse sera tout bonnement installée sur Terre.

Mais cette dernière tourne sur elle-même, tout en tournant autour du Soleil, ces deux mouvements s'effectuant dans le même sens. La vitesse de translation est beaucoup plus importante (de l'ordre de 30 km/s) que celle de rotation. Ainsi Paris, à minuit, se déplace vers l'est à la vitesse de 30 km/s par rapport au référentiel de Copernic.

Imaginons alors les expériences suivantes. Il est minuit, un physicien installé au bois de Boulogne envoie un signal lumineux. Un autre physicien est installé au bois de Vincennes; s'il peut mesurer la vitesse de la lumière qu'il reçoit, nous nous attendons à ce qu'il trouve $c - 30$. En effet, Paris et le bois de Vincennes fuient vers l'est devant le rayon lumineux à la vitesse de 30 km/s.

En revanche, si le physicien du bois de Vincennes envoie un signal lumineux, le physicien du bois de Boulogne, qui vient au-devant de lui avec la vitesse de 30 km/s, devrait trouver une vitesse de $c + 30$. Il semble donc que la lumière ne mette pas le même temps pour le trajet aller que pour le trajet retour. Continuons nos expériences; les physiciens sont maintenant à Sceaux et à Saint-Denis et correspondent par signaux lumineux. Cette fois-ci la lumière aura la vitesse c, mais le rayon lumineux se déplacera latéralement, comme un avion qui vole par vent de travers.

Les expériences que nous venons d'imaginer ne peuvent être précises; les variations de temps de parcours sont de l'ordre (1) du cent-millionième de seconde. Le grand mérite de Michelson est d'avoir conçu une expérience analogue réalisable en laboratoire.

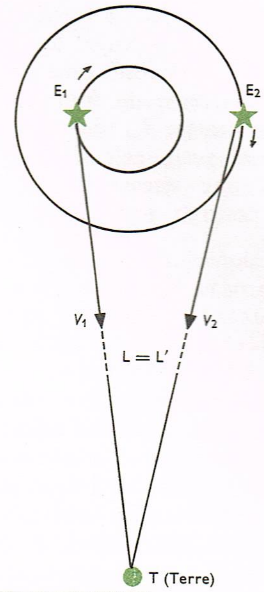
Un rayon lumineux issu d'une source S disposée dans une salle tombe sur une lame de verre A dont la première face est légèrement argentée; cette lame réfléchit une partie de la lumière et laisse passer l'autre partie. Après réflexion sur les miroirs M_1 et M_2 (tels que $OM_1 = OM_2$) placés sur deux bras rectangulaires, on obtient deux rayons lumineux qui ont parcouru l'un le chemin SOM_1OL, l'autre le chemin SOM_2OL. L'analogie avec les expériences imaginées aux environs de Paris est flagrante. Un calcul simple montrerait, en appliquant les hypothèses précédentes, que le temps mis par la lumière pour parcourir le chemin SOM_2OL n'est pas le même que celui mis par la lumière pour parcourir le chemin SOM_1OL.

En inclinant légèrement le miroir M_2, on observe dans la lunette L des franges d'interférence analogues à celles qui sont observées avec un coin d'air.

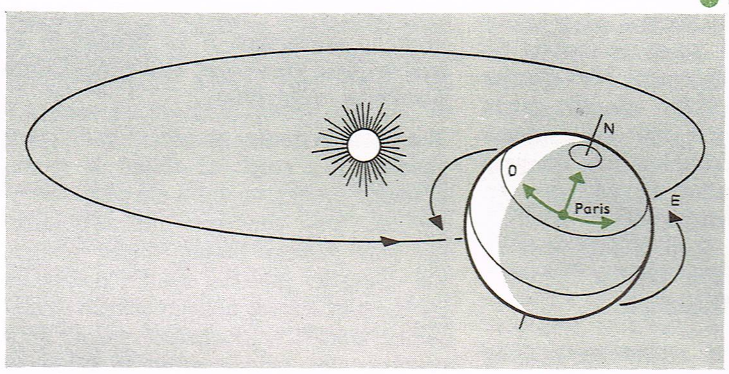
L'expérience décisive est alors la suivante : le montage précédent est placé sur une plate-forme mobile qui peut tourner autour de O, et Michelson fait tourner cette plate-forme d'un quart de tour.

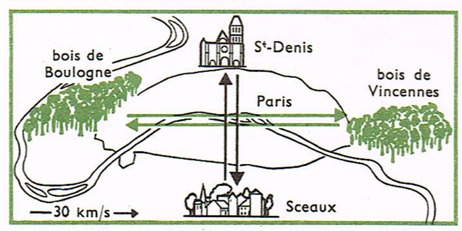
(1) En admettant une distance bois de Boulogne - bois de Vincennes de 12 km :

$$t = \frac{12}{300\,000 - 30} - \frac{12}{300\,000 + 30} \simeq \frac{12.60}{(300\,000)^2} = \frac{720}{9.10^{10}}\,\text{s} \simeq \frac{1}{10^8}\,\text{s}.$$

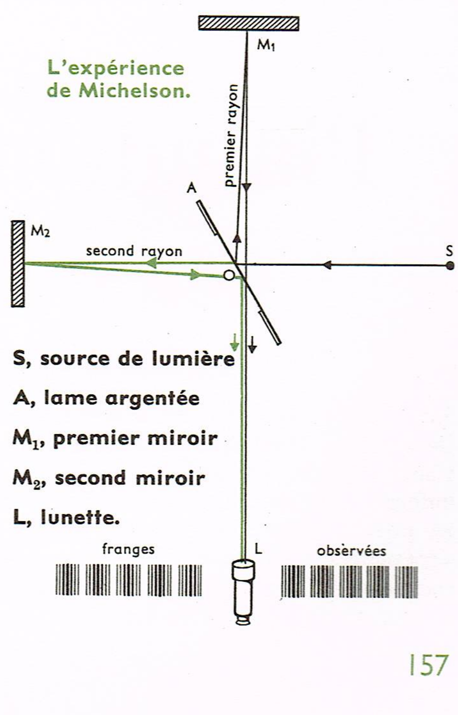


Les rayons lumineux issus des étoiles E_1 et E_2 mettent pour parvenir à la Terre (T) les temps $\Delta t_1 = \frac{L}{v_1}$ et $\Delta t_2 = \frac{L}{v_2}$, qui seront différents si les vitesses de propagation v_1 et v_2 le sont. Les retards considérables (L est très grand) conduiraient à des résultats incohérents sur le mouvement des étoiles, ce que l'on ne constate pas : v_1 est donc rigoureusement égal à v_2.





L'expérience de Michelson.



Les temps mis par les rayons lumineux, toujours si l'on applique les hypothèses classiques, ne seront plus les mêmes. Donc le système de franges observées dans la lunette L doit se déplacer. Michelson avait calculé que, par ce procédé, une vitesse de la Terre de 2 km/s pourrait être mise en évidence.

Or, la réponse de l'expérience a été brutale : **jamais Michelson n'a observé le déplacement attendu des franges d'interférence. Quels qu'aient été le lieu, l'altitude, l'époque de l'année où l'expérience a été réalisée, le résultat a toujours été négatif.** L'expérience, faite par Michelson en 1881, a été répétée par Michelson et Morley en 1887, et, dans de meilleures conditions encore, par Morley et Miller en 1904-1905. Le trajet entre la lame et les miroirs avait été porté à 22 m. Pour une vitesse de 30 km/s, on aurait dû observer un déplacement de l'ordre d'une interfrange, valeur énorme, car la précision des mesures était du centième d'interfrange. Inlassablement, d'autres physiciens ont refait l'expérience en perfectionnant le dispositif de Michelson (Kennedy en 1926, Illingworth en 1927, Joos en 1930). Ils ont confirmé les résultats de Michelson.

D'ailleurs, d'autres expériences électromagnétiques ont été imaginées, et elles aussi ont conduit à des résultats négatifs (Trouton et Noble en 1903, Tomashek en 1927, par exemple).

Tout se passe comme si la Terre était immobile, ou son mouvement indétectable; le désaccord entre la théorie et l'expérience est brutal (à moins de revenir aux théories géocentriques faisant de la Terre le centre et le seul objet immobile de l'univers, autour duquel tournent le Soleil et les étoiles!). **La recherche du repère absolu à l'aide des phénomènes électromagnétiques est un échec.**

attitude d'Einstein

De nombreux physiciens du début du XXe siècle, tels Fitzgerald et Lorentz, ont tenté de sauver le repère absolu par de nouvelles hypothèses. Nous ne les décrirons pas, car elles étaient très artificielles et elles ont été abandonnées progressivement.

Einstein en 1905, à l'âge de vingt-cinq ans, a résolument admis les résultats de l'expérience de Michelson et n'a pas hésité à construire **à la place de la mécanique classique une mécanique nouvelle, relativiste.**

Il a rejeté la notion de repère et d'espace absolus. Il a aussi été obligé de rejeter le temps absolu, ce qui a été plus déconcertant. Des concepts qui semblaient évidents avaient besoin d'être révisés. De nouvelles notions, assez paradoxales au premier abord, allaient prendre naissance. Dans les prochaines pages, nous poserons les principes de la relativité et nous en examinerons les conséquences logiques, vérifiées expérimentalement à leur tour. La relativité n'est pas le résultat d'élucubrations de quelques savants, mais une doctrine cohérente, solidement charpentée, et dont les conséquences pratiques sont vérifiées presque à tout instant à l'heure actuelle. Sachons que la mécanique des grands accélérateurs de particules est relativiste; pensons aux piles atomiques qui utilisent l'énergie nucléaire, prévue par Einstein dès 1906.

Permettons-nous, pour terminer, quelques réflexions philosophiques : nous pensons que, même si l'expérience de Michelson n'avait pas été tentée, les doctrines relativistes auraient fini par triompher. Croire à l'existence d'un repère absolu est un reste d'anthropomorphisme. Pour les Anciens, la fixité de la Terre, centre du monde, était un dogme solide. Pour Copernic, le Soleil autour duquel tourne la Terre et les étoiles restaient fixes. Herschell et les astronomes de la seconde partie du XVIIIe siècle se rendirent compte des mouvements propres des étoiles (et du Soleil par la même occasion). L'éther immobile a résisté un moment au XIXe siècle pour être détruit à son tour par les théories relativistes.

En outre, la mécanique classique, en postulant l'existence du temps absolu, admettait implicitement, comme nous le verrons dans un instant, la possibilité d'être informé instantanément d'un phénomène se produisant à des distances considérables. Quel que soit le courrier adopté, une information ne peut être transmise à distance à vitesse infinie. **Il est choquant que l'explosion d'une étoile appartenant à une nébuleuse située à des dizaines de milliers de milliards de kilomètres soit connue de nous à l'instant même où elle s'est produite.**

l'intuition géniale d'Einstein : deux nouveaux principes

Einstein, en 1905, a tiré les conséquences logiques du résultat négatif de l'expérience de Michelson : le voyageur enfermé dans un référentiel galiléen clos ne pouvait mettre en évidence par des expériences de mécanique son mouvement absolu. Les physiciens du XIXe siècle avaient eu l'espoir d'y parvenir par des expériences d'optique ou d'électromagnétisme, espoir qui, finalement, fut déçu. Aucune expérience de physique, quelle qu'elle soit, effectuée à bord d'un référentiel de Galilée, ne peut déceler son mouvement, ou, ce qui revient au même, les lois physiques sont exactement identiques à bord de deux référentiels galiléens quelconques. Einstein a érigé cette proposition en principe, dit « principe de relativité », dont voici l'énoncé :

Les lois des phénomènes physiques, quels qu'ils soient, sont les mêmes dans tous les systèmes en transla-

tion uniforme les uns par rapport aux autres, c'est-à-dire les mêmes dans tous les référentiels de Galilée.

Signalons bien que ce principe est restreint au cas des systèmes de Galilée, au cas des systèmes non accélérés. Einstein, ne voulant pas voir dans les référentiels de Galilée des référentiels privilégiés, ne s'en contentera pas. Il s'évertuera à énoncer des principes plus généraux et aboutira en 1916 à la théorie de la relativité générale.

Un autre fait étonnant résulte de l'expérience de Michelson. A bord d'un référentiel galiléen, quelle que soit la direction de propagation, la vitesse de la lumière dans le vide est constante. Cela choque notre bon sens : que l'on aille au-devant d'un rayon lumineux ou que l'on se déplace dans le même sens, la vitesse de la lumière est la même. Cela peut nous paraître proprement aberrant, mais nous n'y pouvons rien; les expériences de Michelson et de ses successeurs sont là, souveraines. Avec Einstein, il faut se rendre à l'évidence et énoncer le principe de l'invariance de la vitesse de la lumière :

La vitesse de la lumière dans le vide est égale à une constante absolue *c* dans tous les référentiels galiléens, que la source émettrice soit en mouvement ou non et de quelque façon qu'elle se meuve.

Ces deux principes entraînent obligatoirement la négation de l'espace absolu. Nos habitudes de pensée n'en sont pas choquées outre mesure. Les hommes se sont habitués au fait qu'aucun astre de l'univers ne puisse être le centre fixe du monde. Il n'y a pas de repère absolu; baste! ce n'est pas grave. Nous savons bien que les mouvements s'effectuent toujours par rapport à un repère et que nous ne sommes jamais assurés de sa fixité. Depuis leur tendre enfance, les hommes apprennent que la Terre est animée de mouvements compliqués : rotation, translation autour du Soleil, précession, nutation, etc. Que dire du mouvement du Soleil?

Une conséquence plus sérieuse des deux principes est le rejet de la loi habituelle de composition vectorielle des vitesses. L'addition est-elle donc devenue une opération caduque? Que se passe-t-il? Réfléchissons un peu : une vitesse est le quotient d'un espace parcouru par un intervalle de temps. **Ne serait-ce pas la notion de temps qu'il faille incriminer plutôt que la bonne vieille addition?**

C'est un des titres de gloire d'Einstein que d'avoir révisé le concept de temps : il a retiré au vieux dieu Cronos sa suprématie universelle. Pas plus que l'espace absolu, le temps absolu n'existe. Nous possédons maintenant le courrier idéal, la lumière, pour mettre à l'heure nos chronomètres distants de plusieurs milliers de kilomètres, mais, cruelle ironie, ce courrier idéal a détrôné le temps absolu. Comment? La critique serrée de la notion de simultanéité va nous le montrer.

une notion difficile : la simultanéité de deux événements

Dire que sur la place de l'Hôtel-de-Ville l'éclatement d'un pneu d'une voiture et la sonnerie du beffroi sont simultanés (se sont produits au même instant) ne souffre aucune critique.

Mais, comme nous l'avons déjà noté, dire que Charlemagne est né au moment où éclatait une étoile est bien osé. Pour affirmer cela, il faudrait s'assurer que deux horloges, l'une sur Terre, l'autre sur l'étoile (!), sont synchrones (marquent la même heure).

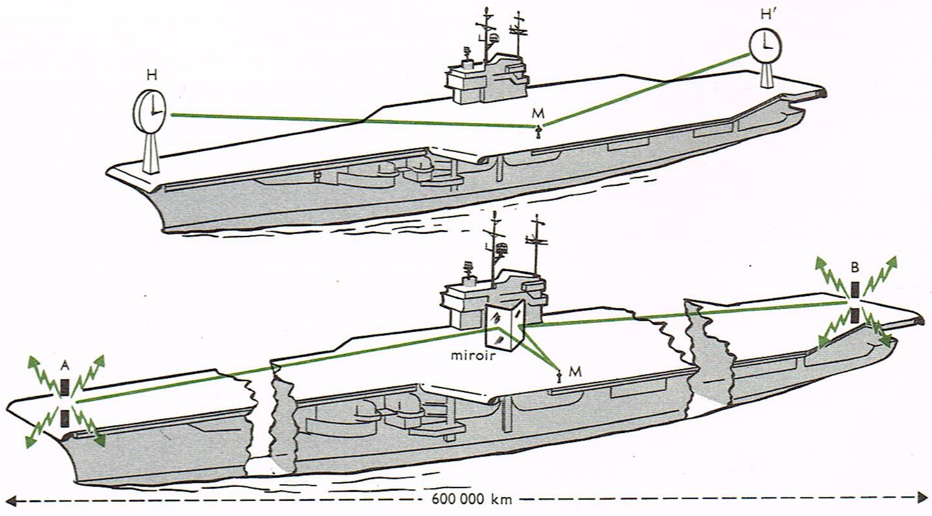
Le problème suivant se pose donc : comment s'assurer de la simultanéité de deux événements se produisant dans des lieux différents? Il faut alors obligatoirement se servir d'un signal. Entre deux astres, entre la Terre et la Lune par exemple, nous avons notamment à notre disposition les ondes électromagnétiques. Elles constitueront d'ailleurs un courrier idéal en vertu du principe de la constance de leur vitesse.

Embarquons de nouveau sur notre porte-avions, qui schématise un référentiel galiléen. A chaque extrémité de la piste se trouve installée une horloge (H et H′); au milieu de la piste, à égale distance de H et de H′, se trouve un marin M. Son rôle est de surveiller que les deux horloges H et H′ marquent bien la même heure. Comment va-t-il faire? Il regarde les deux horloges, et, s'il voit qu'elles marquent la même heure, il est satisfait.

Allongeons démesurément notre référentiel galiléen. Il atteint maintenant 600 000 km de long. A et B sont ses extrémités et M son milieu, où se trouve toujours notre marin hypothétique. En A jaillit une étincelle, et, de même, en B. Si M voit les deux étincelles au même moment, il dira qu'elles ont jailli au même instant. Il dira que les deux étincelles se sont produites simultanément.

Cette définition de la simultanéité de deux événements à distance est bien contrôlable par l'expérience et en complet accord avec le principe de constance de la vitesse de la lumière. Que le référentiel galiléen précédent soit « fixe » ou en mouvement, la lumière mettra une seconde pour venir de A à M aussi bien que pour venir de B à M.

Ainsi, à bord d'un référentiel galiléen, il est possible d'accorder toutes les horloges, de les rendre synchrones, sans déplacer ces horloges. La lumière, courrier idéal, permet de réaliser cet accord.



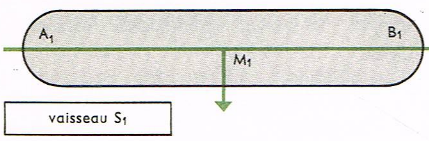
Simultanéité dans un référentiel galiléen.

Tout cela est simple et intuitif. Mais le problème va se compliquer : si nous changeons de référentiel galiléen, deux événements simultanés pour l'un ne le seront pas pour l'autre. Il faut imaginer le croisement de deux référentiels galiléens.

De nouveau, la science-fiction vient à notre secours. Des hommes, intelligents évidemment, et quelque peu physiciens par surcroît, vivent à bord d'un vaisseau spatial d'une centaine de mètres de longueur. Depuis longtemps les réacteurs sont coupés, le navire est loin de toute planète, de tout astre. Les habitants voient toutes les étoiles immobiles. Ils croient, à tort ou à raison, que leur habitacle est au repos dans l'univers.

Ils désirent rendre synchrones des chronomètres disposés en A_1 et B_1 aux extrémités de leur engin.

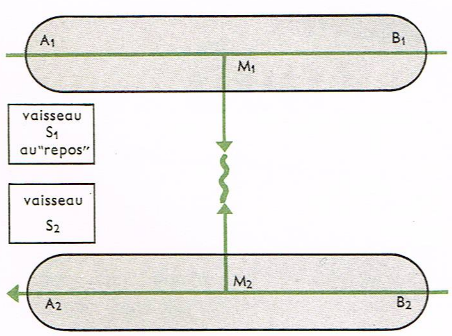
Une étincelle éclate en M_1, situé à égale



distance de A_1 et de B_1 (1). Les observateurs A_1 et B_1 déclencheront leurs chronomètres quand ils verront l'étincelle. La lumière voyageant à la vitesse c par rapport au référentiel S_1 a bien mis le même temps pour parcourir M_1A_1 ou M_1B_1. Les arrivées de la lumière en A_1 et B_1 sont des événements **simultanés** pour nos physiciens.

Par hasard, un second vaisseau spatial S_2, de même longueur, lui aussi habité par des physiciens, croise le premier, et a, par rapport à celui-ci, une vitesse constante v.

Chaque vaisseau porte une pointe en son milieu M_1 ou M_2 et une étincelle



(1) Les physiciens ont vérifié avec un mètre (qui est par définition de longueur constante sur S_1) l'égalité des distances M_1A_1 et M_1B_1.

instantanée jaillit entre ces deux pointes au moment du croisement.

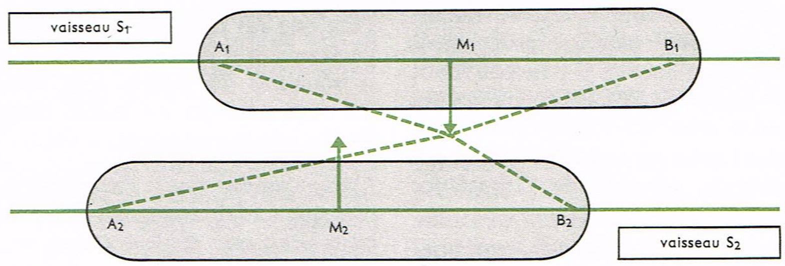
Les quatre physiciens A_1, B_1, A_2, B_2 déclencheront leurs chronomètres quand ils **verront** l'étincelle.

Quel est le raisonnement des physiciens de S_1 ?

a) La lumière partie de M_1 atteint A_1 et B_1 au même moment. Les arrivées du signal en A_1 et B_1 sont des événements simultanés pour les physiciens de S_1;

b) La lumière partie de M_2 à la vitesse **finie** c arrivera en B_2 **avant** d'arriver en B_1, tandis qu'elle arrivera en A_2 **après** être en A_1.

Pour les physiciens de S_1, les arrivées de la lumière en A_2 et B_2 ne **sont pas** des événements **simultanés.**



Allons rendre visite aux physiciens du deuxième vaisseau spatial. Que disent-ils ? « S_2, notre navire, est au repos dans l'univers. S_1 nous croise à la vitesse relative v; c'est bien évident puisqu'il défile devant nous. Les arrivées de la lumière en A_2 et B_2 sont des événements simultanés, tandis que les arrivées de la lumière en A_1 et B_1 ne le sont pas.

Ainsi des événements simultanés pour l'une des équipes ne le sont pas pour l'autre équipe.

Qui a raison? Ni l'un ni l'autre. Aucun des groupes ne peut mettre en évidence son état de repos ou de mouvement. Les deux systèmes galiléens sont rigoureusement équivalents et aucune expérience de physique ne permet de trancher la question.

Les observateurs ont tous tort de vouloir **exporter leur temps.** Les hommes de S_1 définissent un temps valable dans leur référentiel, tandis que les hommes de S_2 définissent un autre temps valable dans leur vaisseau.

L'impossibilité de mettre en évidence l'état de repos absolu ou de mouvement rectiligne uniforme absolu, le principe de la constance de la vitesse de la lumière par rapport à tout repère galiléen **entraînent obligatoirement le rejet de la notion de temps absolu.**

Evidemment, le désaccord ne sera sensible que si la vitesse relative v des deux vaisseaux est de l'ordre de c. Existe-t-il des systèmes animés de très grandes vitesses. Oui; il suffit d'aller à Genève! Le synchrotron du CERN est capable de communiquer des vitesses voisines de celle de la lumière à des particules variées (électrons, protons, antiprotons, etc.).

Ainsi les notions de temps absolu et de simultanéité absolue s'évanouissent. **Deux événements qui ont lieu dans des endroits différents peuvent être simultanés par rapport à un système galiléen S_1 et être séparés par un intervalle de temps fini par rapport à un autre système galiléen S_2.** Nous sommes surpris de cette proposition, et pourtant cette autre ne nous étonne pas :

Deux événements qui ont lieu à des temps différents en un même point d'un système de référence sont séparés par un intervalle d'espace fini par rapport à un autre système. Par exemple, par rapport à l'avion où nous sommes, nous déjeunons et nous dînons au même endroit. Par rapport au sol, le déjeuner et le dîner sont séparés de quelques milliers de kilomètres.

Et pourtant nous n'avons fait qu'intervertir les mots « espace » et « temps » dans les propositions énoncées ci-dessus. Et pourquoi cette impossibilité de définir un temps absolu? **Tout simplement parce qu'il n'existe pas un courrier de vitesse infinie.**

Est-il alors possible qu'un jour on en découvre un, inconnu à l'heure actuelle? Sans vouloir nous lancer dans des prévisions osées et dans des considérations métaphysiques, nous ne le pensons pas. En effet, à tout instant nous serions capables de connaître immédiatement ce qui se passe au même moment dans tout l'univers. Nous aurions presque le don d'ubiquité : quelle prétention!

la transformation de Lorentz

clé mathématique de la relativité

Un retour aux formules classiques de Galilée s'impose tout d'abord. Si un référentiel galiléen S′ (auto) se déplace à la vitesse v par rapport à un autre référentiel galiléen S (Terre), l'abscisse x' (dans le référentiel « auto ») de la pompe à essence M située au bord de la route est liée à l'abscisse x (dans le référentiel « paysage ») de la même pompe par la formule $x' = x - vt$, à laquelle il faut ajouter la deuxième formule, extrêmement simple : $t' = t$. Deux chronomètres, l'un resté au carrefour O, l'autre emporté par l'automobiliste, indiquent la même heure; ce qui veut dire que le temps t mesuré à bord du référentiel « paysage » est le même que le temps t' mesuré à bord du référentiel « auto ».

La transformation de Galilée est en désaccord avec les principes de la relativité développés au chapitre précédent. La loi classique de composition des vitesses est en désaccord avec le principe de constance de la vitesse de la lumière, et nous savons maintenant qu'il faut associer un temps différent à chaque système galiléen, puisque, comme nos deux vaisseaux spatiaux nous l'ont prouvé, il est interdit en relativité d'écrire $t' = t$; le croisement des deux vaisseaux spatiaux l'a prouvé.

Il faut rechercher une autre transformation (d'autres formules liant x, v, t, x', v, t'), qui respecterait les principes de la relativité. Elle s'appelle « transformation de Lorentz », du nom de celui qui l'a découverte vers 1900, avant même Einstein, mais sans en avoir saisi la signification profonde : pour Lorentz, la nouvelle variable temporelle était une simple grandeur mathématique. C'est Einstein qui a eu le mérite de montrer clairement le caractère naturel et physique du temps relativiste.

un convoi relativiste

Le paysage de la page 152 est agréable à regarder. Admirons-le encore, mais cette fois-ci avec des yeux relativistes et non plus avec des yeux « classiques ».

Nous sommes familiarisés avec le référentiel « paysage » auquel est attaché le trièdre Oxyz, et avec le référentiel « auto », qui entraîne son trièdre O′$x'y'z'$. Le poste à essence M est toujours repéré par rapport au carrefour O par son abscisse x, et repéré par rapport au référentiel « auto » par l'abscisse

$$O'M = x'.$$

La question du temps mesuré doit retenir tout particulièrement notre attention. A midi, l'auto est passée avec la vitesse v au carrefour O. Un chronomètre fixé au carrefour O et le chronomètre du conducteur ont alors été déclenchés simultanément. La relativité nous apprend que le chronomètre resté en O ne marquera pas le même temps que celui du conducteur : le chrono de O mesurera le temps t du référentiel « paysage », et le chrono de l'auto le temps t' du référentiel « auto » (en langage mathématique nous dirons que $t = 0$ et $t' = 0$ lorsque O′ coincide avec O).

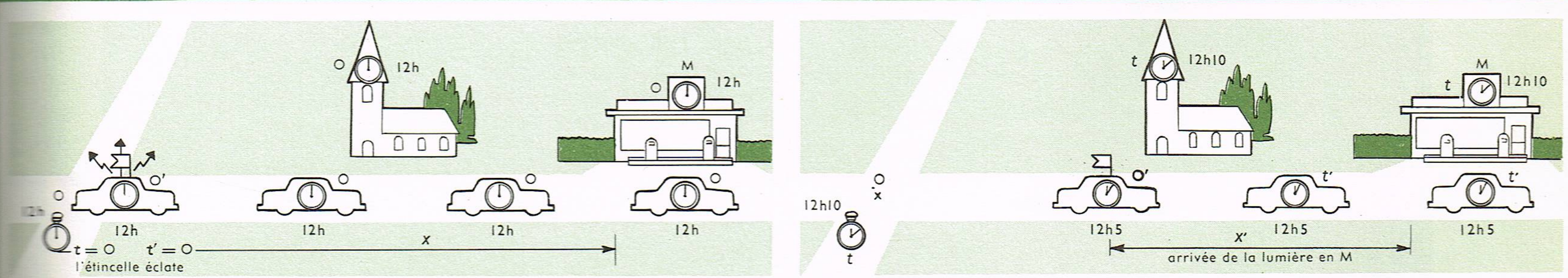
La transformation de Lorentz doit donner les relations liant x' et t' à x et t. Pour découvrir cette transformation, nous avons besoin de notre courrier idéal : le rayon lumineux.

A l'instant précis où l'auto O′ passe au carrefour, un éclair jaillit. Le pompiste du poste M possède un chronomètre indiquant le même temps que le chronomètre situé au carrefour O (le chronomètre du pompiste, celui du carrefour, l'horloge de l'église peuvent être synchronisés puisqu'ils appartiennent au même référentiel « paysage »). Le pompiste note le temps t d'arrivée du signal lumineux; comme la vitesse de la lumière est c, on a $\mathbf{x = ct}$, ou encore, ce qui revient au même, $\mathbf{c^2t^2 - x^2 = 0}$.

Le référentiel « auto » que nous avons utilisé est un peu petit; il est réduit à un seul véhicule. Pour les besoins de la cause, nous allons le compliquer légèrement. Au lieu d'une auto, nous allons en considérer toute une série, toutes équidistantes et toutes roulant à la même vitesse (par exemple, un convoi militaire rigoureusement réglé). Evidemment, l'auto primitive, celle qui passe en O au moment où l'étincelle éclate, doit être distinguée par un fanion puisqu'elle fonctionne comme véhicule « origine O′ ». Tout comme l'auto solitaire, le convoi est un référentiel de Galilée, qui se déplace à la vitesse v par rapport au référentiel « paysage ». Les chronomètres de tous les conducteurs sont à la même heure : l'heure du référentiel « auto ».

Revenons à l'éclair lumineux. Lorsque celui-ci arrive au poste à essence, le conducteur qui est juste à sa hauteur note le temps t' fourni par son chronomètre. Il sait qu'il est à la distance $x' = O'M$ de l'auto d'origine de son convoi, et, comme les chronomètres

On suppose ici que le convoi avance à la vitesse de 60 km/h, et que la vitesse de la lumière est réduite à 300 km/h. Sans cela, quand la lumière arriverait en M, l'auto fanion serait encore pratiquement en O.



du convoi ont été déclenchés au moment du jaillissement de l'éclair, il en déduit que $x' = ct'$, ou encore, ce qui revient au même, que $c^2t'^2 - x'^2 = 0$.

Quelle conclusion tirer de ces observations ? Elle est très simple : si pour le référentiel « paysage » la quantité **$c^2t^2 - x^2$ est nulle, la quantité $c^2t'^2 - x'^2$ associée au référentiel « auto » est nulle aussi.**

Ce double résultat a été obtenu avec le courrier le plus rapide qu'il soit possible d'employer : la lumière. En utilisant un courrier moins rapide, par exemple le son produit par une explosion au temps $t = 0$, au carrefour O, et en reprenant un raisonnement analogue, on montrerait que $c^2t_1^2 - x^2 > 0$ entraîne $c^2t_1'^2 - x^2 > 0$, et réciproquement. On considère cette fois que t_1 et t'_1 sont les temps d'arrivée du son au poste M, comptés l'un par le pompiste, l'autre par le conducteur en coïncidence avec M lors de l'arrivée du son.

à la recherche d'un invariant

Nous avons exploité les principes de la relativité en niant l'existence d'un temps absolu, et en utilisant la constance de la vitesse de la lumière. Pour aller plus loin, faisons un petit raisonnement mathématique.

Pour les deux systèmes galiléens, les quantités $s^2 = c^2t^2 - x^2$ et $s'^2 = c^2t'^2 - x'^2$ sont soit simultanément nulles, soit simultanément positives.

D'où on peut conclure que s'^2 et s^2 sont proportionnels : $s^2 = ks'^2$, et même (les deux référentiels jouant le même rôle) que $s'^2 = s^2$.

Une étude plus générale du mouvement de deux repères galiléens l'un par rapport à l'autre conduirait à l'égalité

$$c^2t^2-(x^2+y^2+z^2)=c^2t'^2-(x'^2+y'^2+z'^2)$$

x, y, z, et x', y', z' étant les coordonnées respectives d'un point M par rapport aux deux référentiels galiléens. Un grand pas a été franchi. Nous venons de découvrir un « invariant » de la relativité. Lorsque l'on passe d'un repère galiléen à un autre repère galiléen, le temps n'est pas invariant; seule la quantité $s^2 = c^2t^2 - l^2$ reste inchangée.

les célèbres formules de Lorentz

Nous verrons ultérieurement que la distance de deux points elle non plus n'est pas invariante. Revenant à l'exemple simple du début, il reste à trouver les formules dites « de Lorentz » qui permettent de passer de t et x à t' et x'. Ces formules doivent être telles que l'égalité fondamentale $c^2t^2 - x^2 = c^2t'^2 - x'^2$ soit satisfaite.

Il est facile de vérifier que (1)

$$x' = \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}} \quad \text{et} \quad t' = \frac{t - vx/c^2}{\sqrt{1 - \frac{v^2}{c^2}}}$$

répondent bien au problème posé.

Pour bien saisir la différence entre la transformation de Lorentz de la mécanique relativiste et la transformation de Galilée de la mécanique classique, écrivons-les face à face :

transformation de Galilée (classique)	transformation de Lorentz (relativiste)
$x' = x - vt$	$x' = \dfrac{x - vt}{\sqrt{1 - \beta^2}}$, où $\beta = \dfrac{v}{c}$ (2);
$t' = t$	$t' = \dfrac{t - vx/c^2}{\sqrt{1 - \beta^2}}$.

Il apparaît tout de suite que si v est petit devant c, les dénominateurs sont pratiquement égaux à 1 et vx négligeable; les formules de Lorentz se ramènent à celles de Galilée, même aux vitesses voisines de $v = 1\,000$ km/s (la vitesse des satellites artificiels n'est que de 10 km/s, et celle de la Terre sur son orbite de 30 km/s) l'écart est faible :

$$x' = 0{,}999\,995\,(x - vt) \sim (x - vt),$$

$$t' = 0{,}999\,995\left(t - \frac{x}{90\,000\,000}\right) \sim t.$$

Pour tous les phénomènes terrestres habituels (à vitesse faible devant celle de la lumière), la mécanique classique constitue une très bonne approximation. Mais ce n'est qu'une approximation, et la portée philosophique des théories relativistes est considérable.

En revanche, pour les particules élémentaires (telles que celles que savent si bien fabriquer les physiciens — à coups de milliards —, ou que l'on observe dans les rayons cosmiques), des vitesses de l'ordre de 270 000 km/s sont courantes, et il ne peut plus être question de négliger les corrections relativistes :

$$x' = 2{,}29\,(x - vt) \neq x - vt,$$
$$t' = 2{,}29\,(t - 0{,}000\,003\,x) \neq t.$$

Les formules — réciproques — permettent de calculer x, t, en fonction de x', t'. Si le système auto S' est animé de la vitesse v par rapport au système paysage S, réciproquement le système S est animé de la vitesse $-v$ par rapport au système S'. Pour les trouver, il suffit de changer v en $-v$.

un premier paradoxe : la contraction des longueurs

Le dessin de la page 164 représente un wagon-laboratoire de la S. N. C. F. en mouvement rectiligne uniforme de vitesse v. Evidemment, le wagon est le référentiel galiléen S', et le paysage devant lequel il défile, le référentiel galiléen S.

Quelle est la longueur de ce wagon ? Il y a deux manières de la mesurer :

1° Les techniciens travaillant dans le wagon peuvent se servir d'un mètre et effectuer leurs mesures à l'intérieur du wagon. Ce seront des observateurs liés au système galiléen « wagon »;

2° Un cheminot d'une gare peut photographier le wagon en marche. S'il connaît l'échelle exacte de la photo, il peut lui aussi connaître la longueur du wagon en mesurant les dimensions de son image sur la pellicule. Le cheminot est un observateur lié au système galiléen « paysage ».

Le bon sens nous dit qu'évidemment la longueur l' du wagon mesurée par les techniciens et la longueur l mesurée par le cheminot fixe sont égales.

La mécanique classique donne d'ailleurs la même réponse. Par rapport à la voie, les abscisses des extrémités A et B du wagon sont x_A et x_B, et la longueur $l = x_B - x_A$ est la longueur photographiée. Les techniciens du wagon prennent, eux, comme abscisses, x'_A et x'_B, c'est-à-dire les distances de A et B à l'origine O', et la longueur qu'ils mesurent est $l' = x'_B - x'_A$. Il suffit d'appliquer la transformation de Galilée ($x' = x - vt$) pour trouver

$$l' = x'_B - x'_A = (x_B - vt) - (x_A - vt) = x_B - x_A = l.$$

---

(1) $c^2t'^2 - x'^2 = \dfrac{c^2\,(t - vx/c^2)^2}{\alpha^2} - \dfrac{(x - vt)^2}{\alpha^2}$, ou $\alpha^2 = 1 - \dfrac{v^2}{c^2}$

$c^2t'^2 - x'^2 = \dfrac{c^2(1 - v^2/c^2)\,t^2 - x^2\,(1 - v^2/c^2)}{\alpha^2} = c^2t^2 - x^2$

par définition même de α.

(2) Par la suite, nous écrirons toujours, pour simplifier l'écriture : $\beta = v/c$ (bêta). En somme, cette convention d'écriture revient à choisir comme unité de vitesse la vitesse de la lumière !

L'EFFET CERENKOV

La lueur violacée que l'on voit autour du cœur des réacteurs du type piscine est due à l'effet Cerenkov (du nom d'un physicien russe à qui ces travaux ont valu un prix Nobel). Ce phénomène s'observe dès qu'une source intense de rayonnements radio-actifs est plongée dans un liquide transparent : on peut le comprendre facilement de façon qualitative.

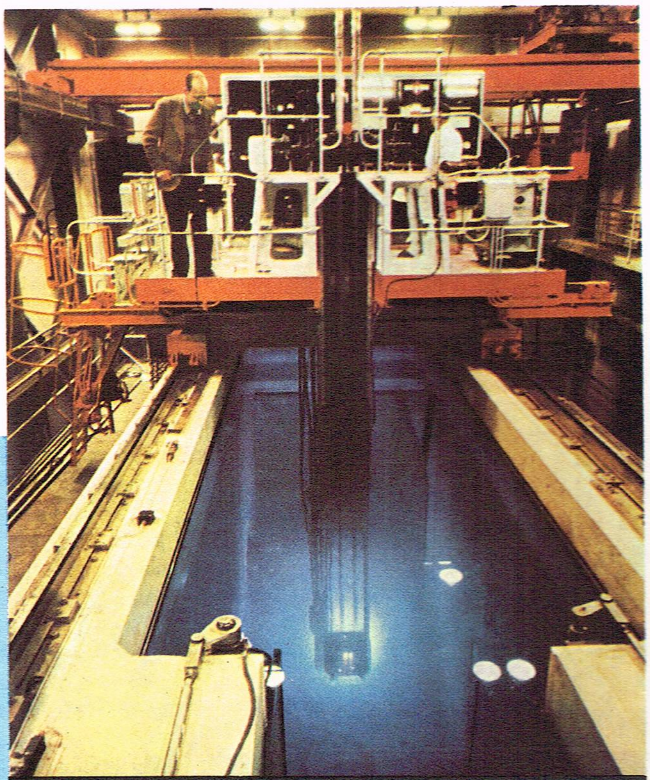
Supposons que des électrons de radio-activité β ou encore des paires d'électrons créées par des rayons γ de grande énergie soient libérés dans l'air : ils sont animés d'une vitesse considérable, qui est presque égale à la vitesse de la lumière dans l'air ou dans le vide. On sait en effet que, pour des énergies de quelques millions d'électron-volts, les électrons sont très fortement relativistes et leur vitesse est déjà proche de sa limite, qui est précisément celle de la lumière.
Imaginons maintenant qu'on fasse pénétrer ces électrons dans un milieu liquide ou solide transparent. Comment d'abord se propage la *lumière* dans un tel milieu? On sait bien qu'elle va moins vite que dans le vide, dans un rapport qui est précisément la valeur du classique indice de réfraction *n*. On a la relation simple suivante :

$$\frac{\text{vitesse de la lumière dans l'eau}}{\text{vitesse de la lumière dans le vide}} = \frac{1}{n}, \text{ soit environ } 0{,}7.$$

Ainsi, l'électron qui entre dans le liquide *avec toute sa vitesse* s'y déplacera *plus vite* que ne pourra le faire la lumière *dans ce même liquide* (alors que dans l'air ou le vide il se déplaçait moins vite que la lumière) : il se produira une onde de choc lumineuse, assez semblable à l'onde de choc acoustique produite par un avion quand sa vitesse devient supérieure à celle des vibrations sonores dans le milieu dans lequel il se déplace.

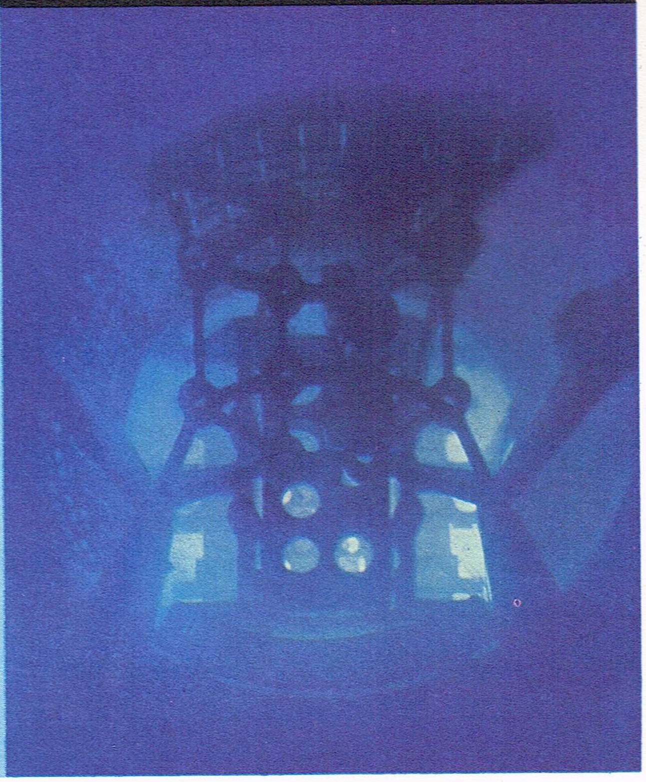
L'onde de choc électromagnétique fait perdre progressivement à l'électron rapide une petite partie de son énergie, et elle apparaît sous forme lumineuse, concentrée dans un cône dont l'axe est la direction de l'électron et dont l'angle d'ouverture est directement relié à la vitesse : plus l'électron va vite, plus le cône est fermé.
Tout cela se calcule et se vérifie fort bien.
Dans l'eau de la pile-piscine, les électrons rapides se déplacent en tous sens, et la lueur que l'on voit représente l'ensemble de ce qui est émis sous forme de lumière visible; c'est un phénomène macroscopique correspondant à un nombre immense de photons Cerenkov.

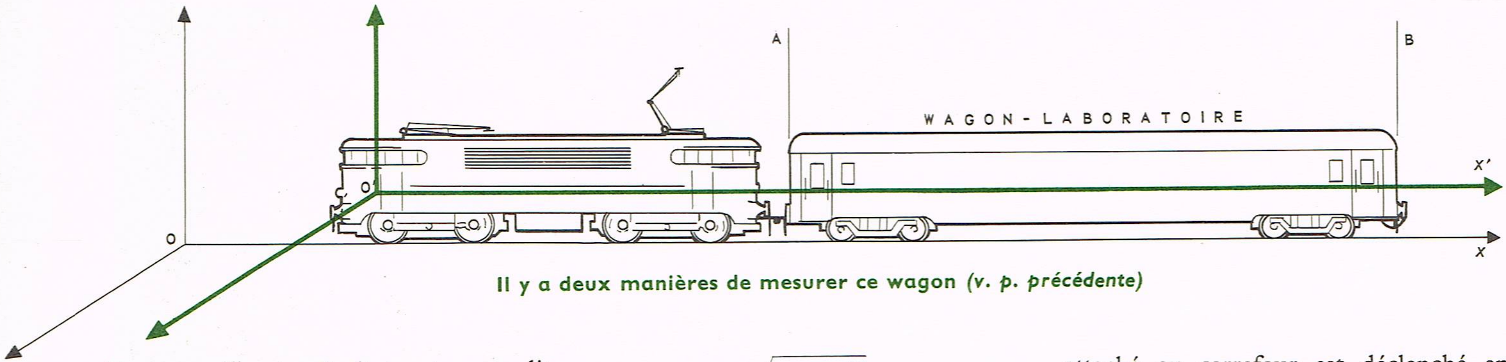
Mais on peut utiliser cet effet de façon extraordinairement raffinée, pour détecter le passage d'une seule particule et même en mesurer la vitesse : si l'on connaît sa direction, et si l'on place sur son passage un liquide ou un plastique transparents d'indice de réfraction connu, le rayonnement Cerenkov de cette particule, c'est-à-dire l'onde de choc lumineuse, sera émis sous un angle *parfaitement défini*, et l'on pourra le recevoir sur une cellule photomultiplicatrice bien placée, qui détectera ainsi le passage de la particule. Bien mieux, si un faisceau contient des particules dont les vitesses sont très grandes sans toutefois être toutes les mêmes, on pourra sélectionner les plus rapides ou les plus lentes et connaître parfaitement leur vitesse, en sélectionnant les angles d'émission des ondes Cerenkov, ce qui est facile avec des dispositifs simples d'optique classiques. Aussi le compteur Cerenkov est-il utilisé dans la physique des hautes énergies, en association avec les faisceaux des grands accélérateurs (faisceaux de mésons, d'antiprotons, etc.) [voir le chapitre sur le CERN]. Ce ne sont plus seulement des électrons qui sont alors intéressés, comme dans le cas de la pile-piscine, ce sont toutes les particules fondamentales possibles auxquelles les grands synchrotrons communiquent des vitesses considérables.



Réacteur-piscine du centre atomique de Harwell, en Grande-Bretagne. *(Doc. Crown.)*

Au centre américain d'Oak Ridge, le fond de la piscine d'un réacteur est éclairé par l'effet Cerenkov. *(Doc. Union Carbide.)*





Il y a deux manières de mesurer ce wagon (v. p. précédente)

En revanche, ce résultat est contredit par la théorie de la relativité. Appliquons, en effet, la transformation de Lorentz, et non plus celle de Galilée. Il reste toujours exact que

$$l = x_B - x_A \quad \text{et} \quad l' = x'_B - x'_A,$$

mais alors

$$l' = x'_B - x'_A = \frac{x_B - vt}{\sqrt{1-\beta^2}} - \frac{x_A - vt}{\sqrt{1-\beta^2}}$$

$$= \frac{x_B - x_A}{\sqrt{1-\beta^2}} = \frac{l}{\sqrt{1-\beta^2}}, \text{ donc } \boldsymbol{l' > l}.$$

Le wagon est plus court pour le cheminot devant lequel passe le wagon que pour les techniciens qui sont dedans.

Le plus étonnant de l'affaire, c'est encore que le phénomène soit réciproque. Toute règle immobile devant laquelle passerait le wagon apparaîtra plus courte (et non plus longue, comme on pourrait s'y attendre) vue du wagon que vue du sol. Il suffit, pour s'en assurer, de refaire le raisonnement précédent. Seul compte le mouvement relatif des deux référentiels, et la formule précédente signifie que la valeur qui résulte de la mesure d'une règle est maximale dans le système de Galilée qui lui est lié (système propre de la règle). Dans tout autre système galiléen, la mesure fait apparaître une contraction dans le rapport

$$\sqrt{1-\beta^2}.$$

Evidemment, un tel phénomène est imperceptible aux petites vitesses : si le wagon roule à la vitesse de 80 km/h, sa longueur se trouve réduite par un facteur de 0,999 999 999 999 99. Ce qui correspond à une diminution de l'ordre du diamètre d'un atome! Une fusée interplanétaire de 100 m de long voguant à 400 000 km/h est raccourcie de 0,01 mm, ce qui est déjà important, mais encore bien faible.

un second paradoxe : la dilatation du temps

Les temps mesurés dans deux référentiels galiléens sont différents. La critique serrée de la notion de simultanéité a établi ce fait important. Les formules de Lorentz montrent bien, d'autre part, que t' est différent de t. Il est intéressant maintenant de trouver une relation précise entre les deux temps associés aux deux repères.

Une fois de plus (v. p. *ci-contre*) examinons notre référentiel « paysage » et notre référentiel « auto ». Lorsque l'auto franchit le carrefour O, un chronomètre attaché au carrefour est déclenché en même temps que celui du conducteur.

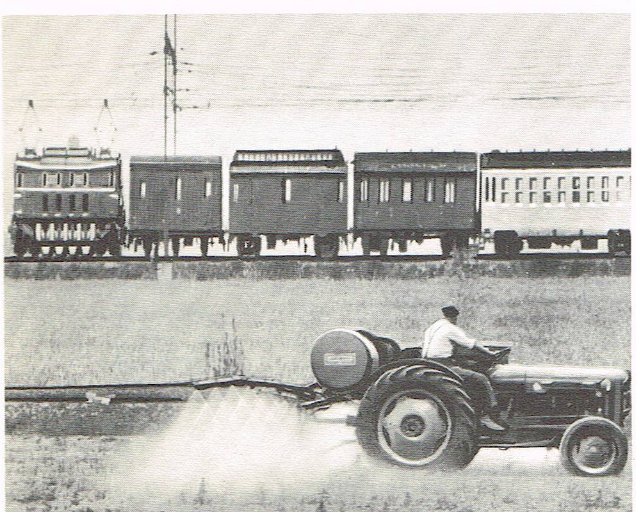
Le chronomètre du conducteur installé à l'origine O′ du référentiel « auto » mesure le temps t'_1, tandis que le chronomètre fixé au carrefour O mesure le temps t_1. Ecrivons la formule de Lorentz.

$$t = \frac{t' + \frac{vx'}{c^2}}{\sqrt{1-\beta^2}}.$$

Dans cette formule, x' est l'abscisse du chronomètre du conducteur. Puisque cette montre est en O′ évidemment, $x' = 0$, et la formule donne

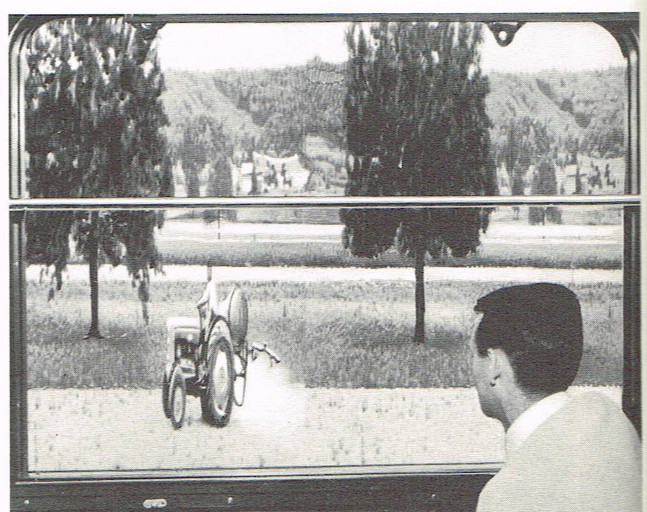
$$t_1 = \frac{t'_1}{\sqrt{1-\beta^2}}.$$

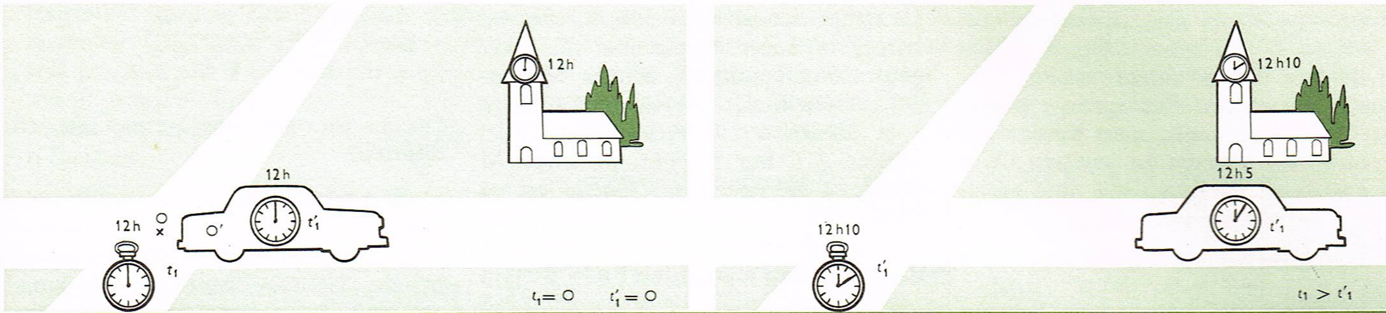
t_1 est plus grand que t'_1 puisque $\sqrt{1-\beta^2}$ est inférieur à l'unité.
Les phénomènes se déroulant à bord du référentiel « auto » sont ralentis dans le rapport $\sqrt{1-\beta^2}$ pour les observateurs du référentiel « paysage ». **Là aussi, le plus étonnant est la réciprocité du ralentissement du temps.** Il est à noter d'ailleurs que cette réciprocité est en accord avec le principe de relativité : rien ne doit distinguer deux repères galiléens en mouvement l'un par rapport à l'autre.



Comment le conducteur du tracteur verrait passer un train relativiste si la vitesse de ce dernier approchait l'ordre de grandeur de celui de la lumière...

... et comment le voyageur verrait le tracteur et le paysage.





Les phénomènes se déroulant à bord du référentiel « auto » sont ralentis pour les observateurs du référentiel « paysage ».

En effet, supposons maintenant que le chronomètre du carrefour indique le temps t_2; celui de l'automobiliste indiquera le temps t'_2. La formule de Lorentz

$$t' = \frac{t - \frac{vx}{c^2}}{\sqrt{1 - \beta^2}}$$

donne $t'_2 = \frac{t_2}{\sqrt{1 - \beta^2}}$, puisque au carrefour O, origine du référentiel « paysage », x est égal à 0. Il est évident que t'_2 est plus grand que t_2. Tous les phénomènes se déroulant à bord du référentiel « paysage » sont ralentis dans le rapport $\sqrt{1 - \beta^2}$ pour l'automobiliste.

Il est superflu de rappeler que tous ces phénomènes étonnants (contraction des longueurs, dilatation des temps) ne sont appréciables que si β est voisin de 1, que si la vitesse v est de l'ordre de 300 000 km/s. Les exemples choisis (référentiel « paysage », référentiel « auto », référentiel « porte-avions », etc.) ont permis de comprendre ces paradoxes; ils n'avaient pas l'ambition de décrire des expériences capables de les mettre en évidence. Exploitons ce que nous venons de découvrir : la durée d'un phénomène physique se produisant à bord d'un repère galiléen est toujours trouvée plus courte par un observateur lié au repère que par un observateur associé à un autre système.

un temps privilégié : le temps propre

En d'autres termes, il y a une infinité de mesures de la durée d'un phénomène, puisqu'il y a une infinité de référentiels galiléens d'où l'on peut étudier ce phénomène. Parmi les nombreuses valeurs de cette durée, il en est une qui est plus petite que toutes les autres : c'est la valeur trouvée quand on étudie le phénomène dans le repère galiléen auquel il est lié. Cette durée minimale est appelée la **durée propre du phénomène.** Les horloges emportées par un repère galiléen mesurent le temps propre du système galiléen. Cette notion de temps propre est essentielle, et elle nous rassure au moment même où nous étions submergés par cette infinité de valeurs possibles du temps. Nous venons de découvrir une notion simple, intuitive, qui nous est attachée. Le temps absolu de la mécanique classique était un concept lointain, insaisissable; **le temps propre est le temps « que nous emportons avec nous » : il nous est indissolublement associé.**

Lorsque les physiciens parlent de la durée de vie d'une particule (par exemple, la durée de vie d'un méson μ est de l'ordre de 10^{-6} s), ils veulent dire « durée de vie propre ». En faisant un abus de langage, nous dirons que c'est le temps « senti » par la particule.

Continuons à approfondir cette notion si passionnante. Les physiciens, toujours aussi épris de l'alphabet grec, ont l'habitude de désigner le temps propre par la lettre τ.

La relation entre le temps t, relatif à un repère galiléen animé de la vitesse $v = \beta c$ par rapport à un phénomène physique, et le temps propre τ de ce même phénomène s'écrit :

$\boldsymbol{\tau = t\sqrt{1 - \beta^2}}$, d'où il résulte que

$$\tau^2 = t^2(1 - \beta^2) = t^2 - \frac{v^2t^2}{c^2} = \frac{c^2t^2 - x^2}{c^2} = \frac{s^2}{c^2},$$

$s^2 = c^2t^2 - x^2$ étant l'invariant fondamental associé au point où se trouve l'horloge mesurant le temps propre.

La quantité s^2 étant constante quel que soit le repère galiléen adopté, il en est de même du temps propre.

Le temps propre est un invariant par rapport à toute transformation de Lorentz.

C'est une excellente propriété, et bien rassurante : le temps propre joue effectivement un très grand rôle en relativité restreinte.

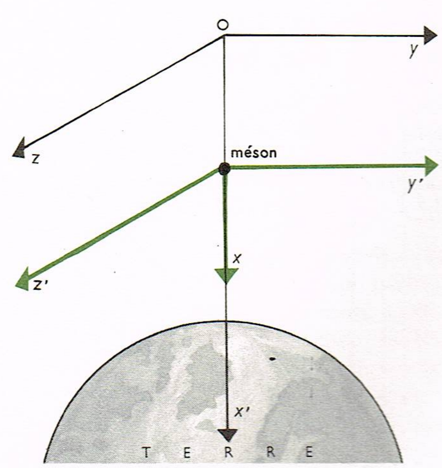
ce que vivent les mésons

Les mésons μ sont des particules de masse 200 fois plus grande que celle de l'électron, qui ont été décelées dans le rayonnement cosmique. Les mésons μ se désintègrent après une vie moyenne τ en un électron et un neutrino.

Les physiciens ont pu mesurer la vie moyenne τ_0 des mésons au repos, en notant l'instant de leur entrée dans une masse de métal et celui où ils s'y arrêtent; et ils ont trouvé $\tau_0 = 2{,}2.10^{-6}$ s.

Or, ces mésons décelés dans le rayonnement cosmique peuvent parcourir avec une vitesse voisine de celle de la lumière plusieurs kilomètres avant de se désintégrer. Pourtant, si leur vie moyenne était τ_0, ils ne pourraient guère parcourir plus de
$L - c\,\tau_0 = (3.10^8).\ (2{,}2.10^{-6}) = 660$ m.

Le désaccord est flagrant. C'est la relativité qui va nous donner la clef de la contradiction apparente : τ_0 est la durée de vie du méson dans son système



propre (système par rapport auquel le méson est au repos). Or, le méson se déplace à une vitesse voisine de c dans l'atmosphère. Son système propre est $O'x'y'z'$, et nous, nous faisons nos mesures par rapport au système $Oxyz$ lié à la Terre. Le temps que nous mesurons n'est pas τ_0, mais τ, tel que

$$\tau = \frac{\tau_0}{\sqrt{1 - \beta^2}}.$$

Le parcours du méson dans l'air est donné par :

$$L \simeq \tau_0 = \frac{\tau_0\, c}{\sqrt{1 - \beta^2}}.$$

$\sqrt{1 - \beta^2}$ est très faible; donc ce parcours L est de beaucoup supérieur à 660 m, ce qu'ont vérifié très exactement, en 1941, Rossi et Hall.

Un atome qui rayonne est une horloge minuscule. Supposons qu'au repos il rayonne une radiation de longueur d'onde λ. Si l'atome se déplace par rapport à l'observateur avec la vitesse v, l'observateur note une variation de longueur d'onde $\Delta\lambda$ telle que $\Delta\lambda = \lambda \frac{v}{c}$. C'est l'effet Doppler-Fizeau classique. Mais ici, du fait de son mouvement, la petite horloge atomique est ralentie, et là encore le facteur $\sqrt{1 - \beta^2}$ intervient : le décalage observé $\Delta\lambda$ doit être plus grand que celui prévu par l'effet Doppler-Fizeau classique.

Ce décalage a été mis en évidence par les physiciens Ives et Stilwell de 1938 à 1941, dans des expériences délicates dont la description précise se trouve dans des ouvrages spécialisés. (Notons à ce propos un fait amusant : Ives était antirelativiste, et il avait imaginé ces expériences pour appuyer sa doctrine. Il a été convaincu bien malgré lui et l'a reconnu en toute honnêteté scientifique.)

où deux et deux ne font plus quatre

Gagnons le bord de la mer et intéressons-nous de nouveau à l'aéronavale. Le temps est au beau fixe; la mer est calme, la houle négligeable. La mer pourra être considérée comme référentiel galiléen S. Un porte-avions navigue à la vitesse v (ce sera le référentiel galiléen S'). Sur le pont, un avion M décolle à la vitesse u'. Quelle est sa vitesse par rapport à la mer? Sans réfléchir, et en accord avec la mécanique classique, nous répondrons : $u = u' + v$.

Un autre exemple. Nous sommes sur un navire interplanétaire avançant à la vitesse de 200 000 km/s par rapport à la Terre (nous faisons là de la haute anticipation, mais qu'importe!). Par un procédé digne de J. Verne, de ce vaisseau est lancé un boulet à la vitesse de 200 000 km/s. La vitesse résultante est de 200 000 + 200 000 = 400 000 km/s par rapport à la Terre.

C'est la mécanique classique qui nous enseigne cela. Cette loi de composition des vitesses est une conséquence de la transformation de Galilée, alors qu'en relativité nous admettons la transformation de Lorentz.

Il est bien évident que la transformation de Lorentz doit conduire à une loi de composition des vitesses différente de la précédente, et pour une double raison : x n'est plus égal à $x' + vt$, et les temps t et t' pour les deux référentiels ne sont plus les mêmes. Les deux formules précédentes permettent d'établir la formule relativiste de composition des vitesses (1) :

(1) Supposons que la vitesse du mobile par rapport au repère S soit constante : $u = \frac{x}{t}$;

$$x = \frac{x' + vt'}{\sqrt{1 - \beta^2}}; \qquad t = \frac{t' + \frac{v}{c^2}x'}{\sqrt{1 - \beta^2}}; \text{ d'où}$$

$$u = \frac{x}{t} = \frac{x' + vt'}{t' + \frac{v}{c^2}x'} = \frac{\frac{x'}{t'} + v}{1 + \frac{v}{c^2}\frac{x'}{t'}}, = \frac{u' + v}{1 + \frac{u'v}{c^2}},$$

puisque par rapport au référentiel S' la vitesse est $u' = \frac{x'}{t'}$.

$$u = \frac{u' + v}{1 + \frac{u'v}{c^2}}.$$

Quelles sont les conséquences de cette formule?

a) En premier lieu, elle ne se différencie de la loi classique que si u' et v sont de l'ordre de c. Là encore, pour les vitesses faibles (vitesses jusqu'à 10 000 km/s), on peut écrire sans grande erreur :

$$u = v + u'.$$

b) Supposons maintenant que le point M (M est alors un grain de lumière, un photon) soit animé de la vitesse $u' = c$ par rapport au référentiel S'. Nous sommes curieux de savoir quelle sera la vitesse u du photon par rapport au référentiel S.

$$u = \frac{c + v}{1 + \frac{cv}{c^2}} = \frac{(c + v)\,c^2}{c^2 + cv} = \frac{(c + v)\,c^2}{c\,(c + v)} = c.$$

Eh bien! la vitesse du photon par rapport à S est encore c, et nous énoncerons ce nouveau résultat fondamental :

La vitesse de la lumière c est une vitesse limite qui ne peut en aucun cas être dépassée.

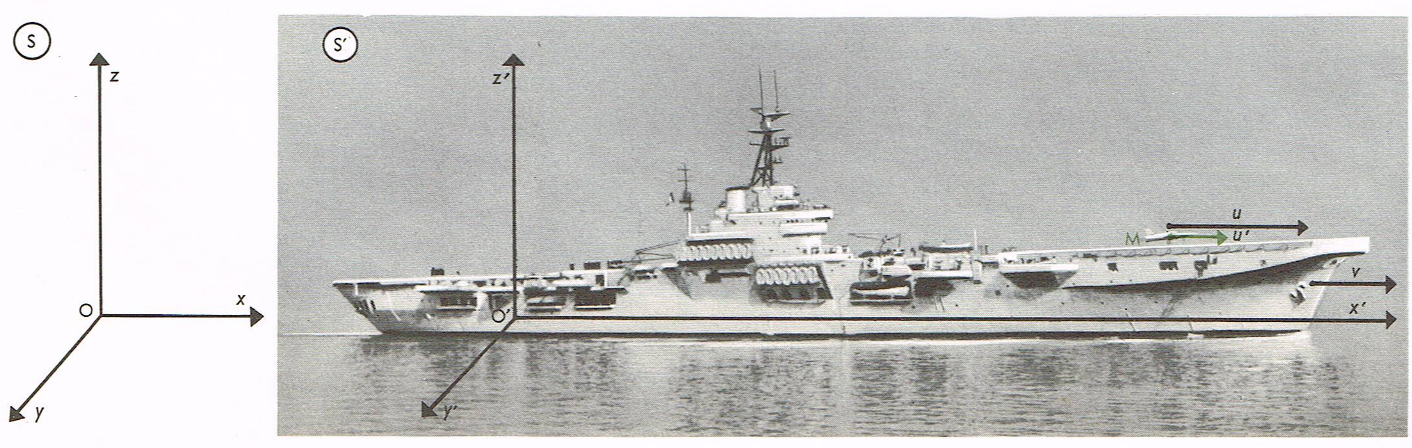
Nous avons donc là la confirmation qu'aucun signal de nature quelconque ne peut dépasser la vitesse de la lumière.

En particulier, nous retrouvons ici l'impossibilité d'une vitesse de propagation infinie, fait sur lequel nous avons déjà insisté.

Reprenons notre exemple numérique précédent du boulet lancé à 200 000 km/s à partir du navire fonçant à 200 000 km/s. Sa vitesse vaut

$$u = \frac{200\,000 + 200\,000}{1 + \frac{200\,000 . 200\,000}{(300\,000)^2}} = 277\,000 \text{ km/s}.$$

Nous trouvons bien, comme il se doit, une vitesse inférieure à la vitesse de la lumière, et non pas 400 000 km/s comme le prévoyait la mécanique classique.



Phot. Stéphane Murat.

Existe-t-il une expérience vérifiant cette formule? Oui, mais si paradoxal que cela paraisse, elle a été faite par le physicien français Fizeau en 1851, cinquante-quatre ans avant les mémoires d'Einstein.

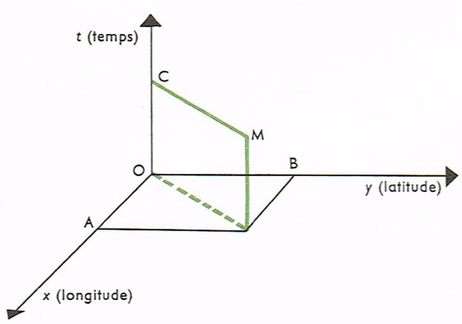
Nous n'entrerons pas dans la description détaillée de cette expérience. Elle consiste à étudier la vitesse de propagation des ondes lumineuses dans de la matière transparente en mouvement (de l'eau). La formule trouvée par Fizeau fait intervenir un coefficient d'entraînement qui ne peut s'expliquer en physique classique que par un entraînement partiel (?) du fameux « éther », interprétation d'autant plus étrange que cet entraînement de l'éther dépendrait de l'indice de réfraction de la substance.

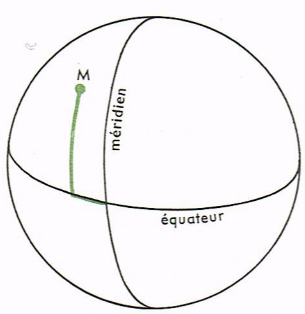
En revanche, en utilisant la formule relativiste de composition des vitesses, on retrouve immédiatement et très simplement le résultat expérimental de Fizeau.

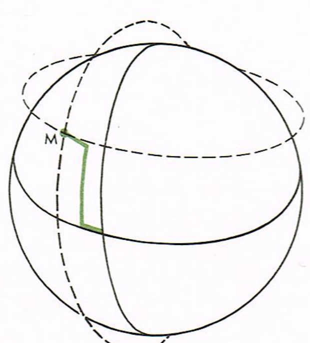
Voici étudiées les principales conséquences de la transformation de Lorentz, donc des principes de la relativité restreinte. Des résultats étranges se sont dévoilés, qui ont été vérifiés par des expériences nombreuses et précises dont nous avons essayé de donner une idée succincte.

Si les conséquences de la théorie relativiste se réduisaient à celles-là, ce serait déjà très important. Mais, jusqu'ici, nous n'avons fait que de la cinématique relativiste. Il va nous falloir aborder la dynamique relativiste, et là les conséquences pratiques seront de portée incalculable. C'est à l'orée de l'ère nucléaire qu'elles nous placent.

l'espace - temps







Des « graphiques » à deux, trois et quatre dimensions sont nécessaires pour repérer respectivement la position M d'un train *(p. 168)*, d'un navire *(ci-dessus, au centre)* et d'un avion *(en bas)*.

les graphiques de chemin de fer

1° Examinons la marche du « Mistral » entre Dijon et Lyon (p. 168). Le graphique correspondant permet à tout instant t de connaître la position du train déterminée par sa distance x de Paris.

Nous nous sommes servis pour cette représentation de deux axes de coordonnées : un axe temporel t, gradué en heures et en minutes, et un axe spatial, gradué en kilomètres.

2° Pouvons-nous maintenant imaginer de même un graphique permettant de repérer à tout instant la position d'un navire au milieu des océans? Le graphique précédent à deux dimensions, l'une spatiale, l'autre temporelle, sera insuffisant. Deux coordonnées sont nécessaires pour repérer un navire sur l'océan : sa longitude et sa latitude. Avec l'axe temporel t, nous voici obligés d'avoir recours à un espace à trois dimensions. Chaque état possible du navire sera représenté par un point M de cet espace.

Le 12 juillet 1962, à 9 h 20, le navire sera par 35° de longitude ouest et 45° de latitude nord. OA représentera 35° de longitude; OB, 45° de latitude; et OC, l'heure et le jour en prenant une origine des temps convenable, par exemple le départ du navire du Havre.

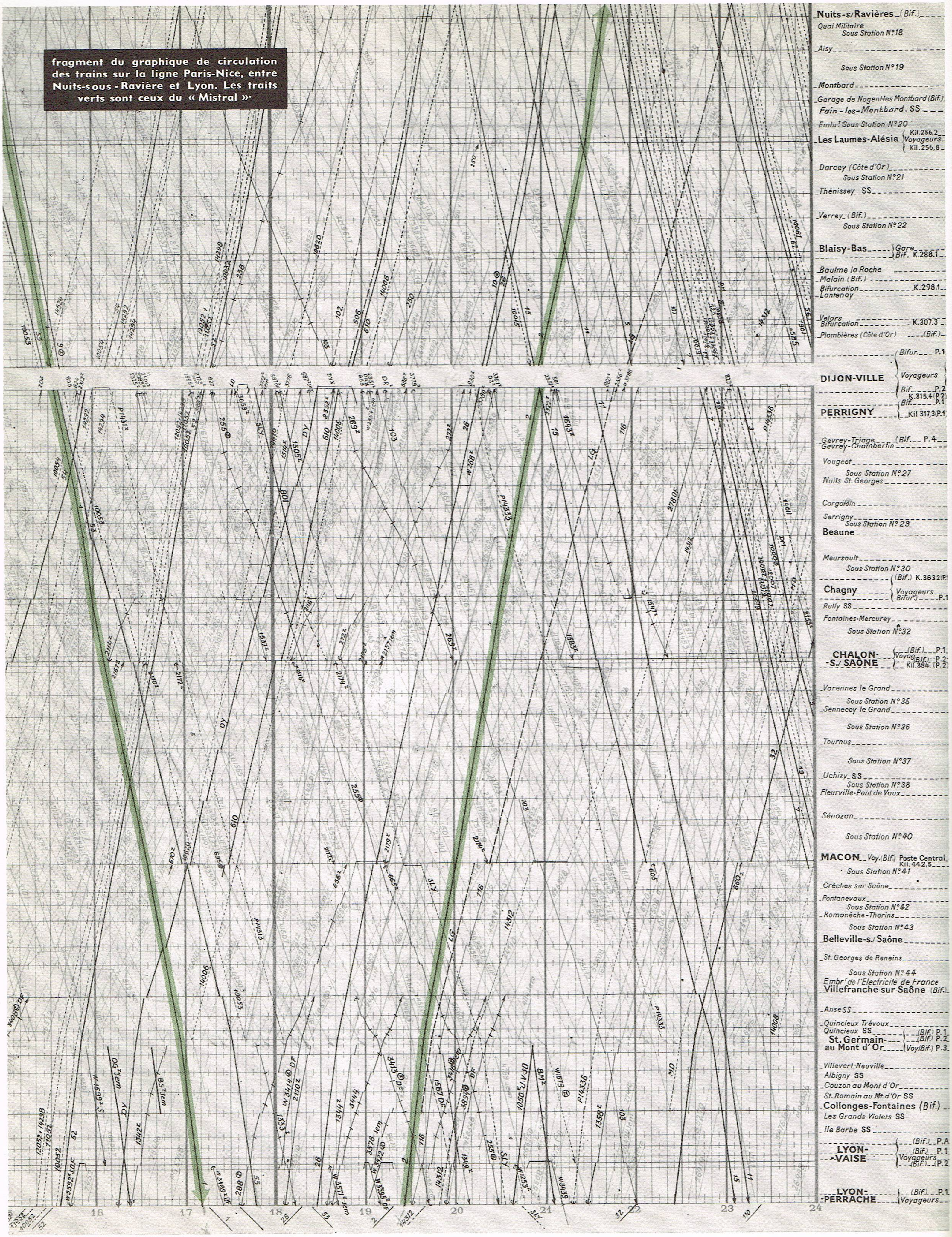
3° Et si maintenant nous imaginions un « graphique » permettant de repérer à tout instant la position exacte d'un avion? Il nous faut trois dimensions spatiales, latitude, longitude, altitude de l'avion, et une dimension temporelle. Nous voici donc obligés d'avoir recours à un espace à quatre dimensions. Cette fois, nous ne pouvons le dessiner sur une feuille de papier, même avec les ressources de la perspective. Mais nous pouvons imaginer un tel espace. Nous admettrons quatre axes de coordonnées, et à tout point de cet espace seront associés quatre nombres : longitude, latitude, altitude, époque de la position de l'avion.

Signalons au passage que les mathématiciens, gens à l'imagination féconde, ont construit des espaces à 4, 5, 6, ... n dimensions. De tels espaces rendent de grands services en mécanique statistique. Ils ont même été jusqu'à étudier les propriétés d'espaces à un nombre infini de dimensions (utilisés par les mathématiciens et les physiciens en théorie quantique des champs). Rassurons-nous; dans notre exposé, nous nous contenterons de quatre dimensions.

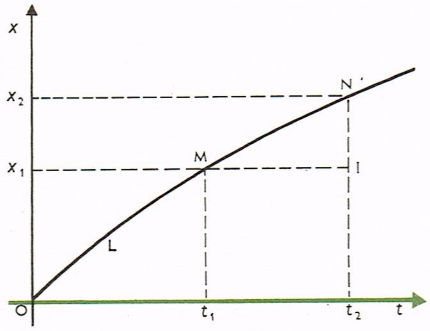
distance de deux points dans un espace donné

Reprenons, par souci de simplicité, un graphique à deux dimensions. La ligne L représente à tout instant t la position x d'un mobile sur une droite.

Le point M a pour coordonnées l'instant t_1 et la distance x_1. Le point N, par contre, a pour coordonnées t_2 et x_2. Peut-on parler de la distance MN des deux points M et N? Oui, répondons-nous, puisque, sur le graphique, en appliquant un double décimètre, nous pouvons mesurer MN.



fragment du graphique de circulation des trains sur la ligne Paris-Nice, entre Nuits-sous-Ravière et Lyon. Les traits verts sont ceux du « Mistral »



Il est vrai que $MN^2 = MI^2 + IN^2$, mais cela n'a guère de sens : peut-on additionner des carrés d'heure et des kilomètres carrés ?

En réalité, cette mesure ne signifie absolument rien. Prouvons-le en ayant recours au théorème de Pythagore. Nous sommes tentés d'écrire $MN^2 = MI^2 + IN^2$.

Eh bien! l'égalité que nous venons d'écrire n'a aucune signification.

En effet, si $MI = t_2 - t_1 = 3$ h, et $IN = x_2 - x_1 = 150$ km, MI^2 vaut 9 heures carrées, et IN^2, 22 500 km². Mais comment procéder à l'addition de MI^2 et de IN^2? On peut additionner entre eux des kilomètres, des kilomètres carrés, des heures, mais que peut bien signifier 9 heures² + 22 500 km²?

Un problème important se pose : quand pouvons-nous donner à la distance MN une signification précise? La réponse est immédiate : il est nécessaire que les axes de coordonnées soient gradués en unités identiques, par exemple en unités de longueur.

Comment passer alors d'un temps à une longueur? Tout simplement en multipliant le temps par une vitesse. Si t est un temps et si v est une vitesse, la quantité $l = vt$ a bien les propriétés (dimensions) requises pour être comparée à une longueur.

Depuis le début de ce chapitre, il semble que nous ayons oublié nos préoccupations relativistes. Il n'en est rien. Quelle vitesse choisir pour convertir le temps en longueur? Autant que possible, il faut choisir une vitesse invariante par rapport au système de coordonnées utilisé. Il est évident que la vitesse c de la lumière va faire notre affaire (1).

Les quatre coordonnées repérant un point M de l'espace-temps seront :

1° x, y, z, coordonnées spatiales;

2° $\lambda = ct$, coordonnée temporelle multipliée par la vitesse de la lumière.

Nous appellerons tout point M de l'espace-temps un « événement ». **Un événement est donc repéré par l'endroit où il s'est passé et par la date à laquelle il a eu lieu.**

distance de deux événements de l'espace-temps

Revenons à la physique classique et rappelons-nous ce que nous avons établi dans l'introduction.

Dans l'espace absolu de la mécanique classique, la distance de deux points M et N est un invariant absolu. Cette distance est indépendante du système de référence choisi. En particulier, si nous prenons des systèmes ayant tous la même origine O, la distance à l'origine est

$$l^2 = OM^2 = x^2 + y^2 + z^2.$$

Si l'on change de repère Oxyz, x, y, z deviennent x', y', z', mais on a toujours $l^2 = OM^2 = x'^2 + y'^2 + z'^2$. La longueur de la règle OM est invariante.

En mécanique classique, il n'y a de temps qu'absolu : quel que soit le système de référence adopté, t est partout le même. La physique classique connaissait donc deux invariants : l (longueur) et t (temps).

Par contre, en relativité, les notions d'espace et de temps absolus s'effondrent. Nous savons que deux observateurs en mouvement l'un par rapport à l'autre ne trouveront pas la même longueur pour une règle donnée. Nous savons que deux observateurs en mouvement l'un par rapport à l'autre posséderont des montres qui n'indiqueront pas la même heure. Donc, en relativité, ni la quantité $l^2 = x^2 + y^2 + z^2$, ni t ne sont invariants quand on change de système galiléen. Mais, fait fondamental qu'il ne faut jamais perdre de vue, la relation

$$\mathbf{s^2 = c^2t^2 - x^2 - y^2 - z^2 = c^2t^2 - l^2}$$

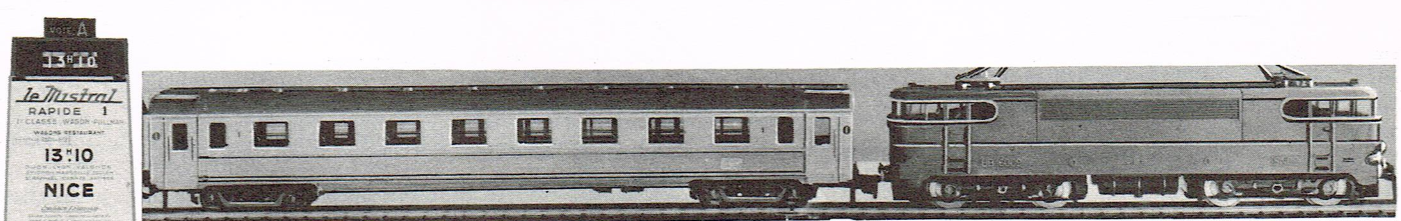
reste invariante quand on passe d'un système galiléen à un autre système galiléen.

Nous dirons encore que la distance spatio-temporelle de deux points de l'espace-temps à quatre dimensions (deux événements) est définie par

$$s^2 = c^2 \Delta t^2 - [\Delta x^2 + \Delta y^2 + \Delta z^2].$$

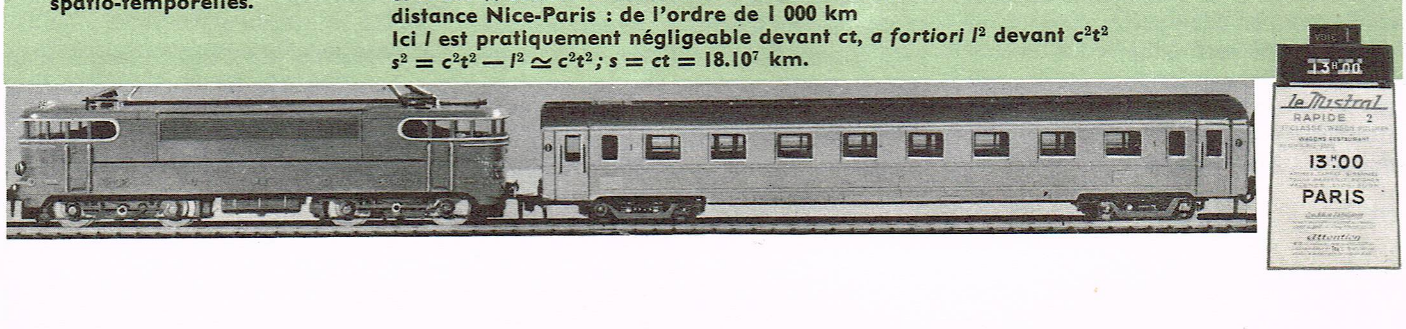
Donc la distance spatio-temporelle d'un événement M de l'espace-temps à un événement origine arbitrairement choisi

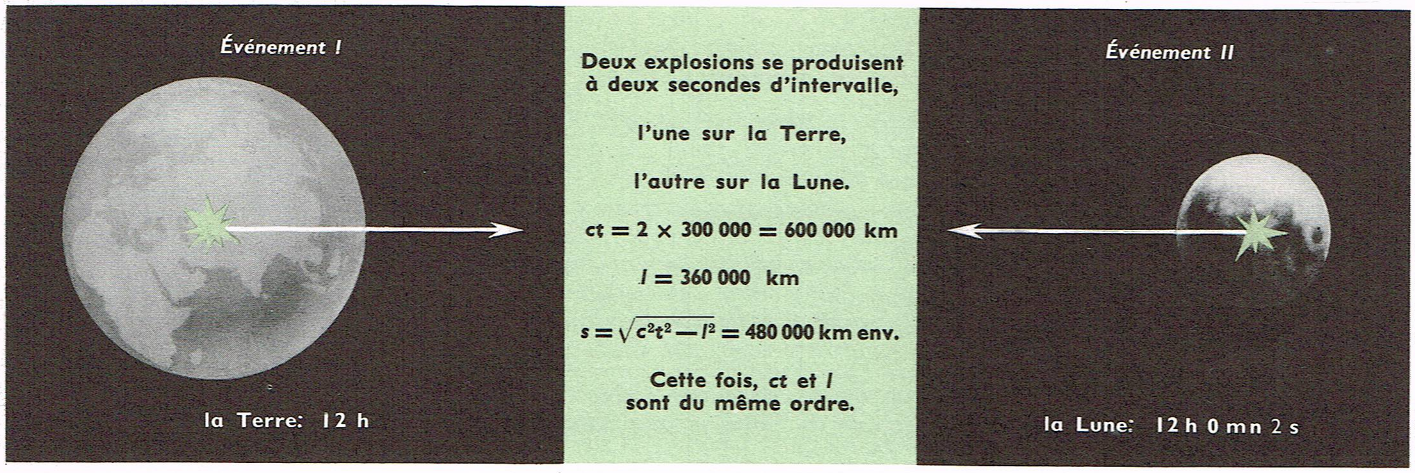
(1) Bien avant les théories relativistes, les astronomes convertissaient les longueurs en temps à l'aide de la vitesse de la lumière. Les distances de la Terre aux étoiles étant par trop grandes, ils utilisèrent l'année de lumière : c'est la distance parcourue par la lumière en un an, soit 946 300 000 000 000 de kilomètres. Évidemment, à l'inverse, le kilomètre-lumière serait un intervalle de temps extrêmement petit, le temps mis par la lumière pour parcourir 1 kilomètre étant 1/300 000 s. Cette unité n'a jamais été utilisée.



Calcul des distances spatio-temporelles.

Événement I : départ du train « Mistral » aller de la gare de Lyon à 13 h 10.
Événement II : départ du train « Mistral » retour de la gare de Nice à 13 h.
$t = 10$ minutes ou 600 secondes
$ct = 600 \times 300\,000 = 18.10^7$ km = 180 000 000 km
distance Nice-Paris : de l'ordre de 1 000 km
Ici l est pratiquement négligeable devant ct, *a fortiori* l^2 devant c^2t^2
$s^2 = c^2t^2 - l^2 \simeq c^2t^2$; $s = ct = 18.10^7$ km.





(origine à la fois de l'espace et du temps) est donnée par
$\mathbf{s^2 = \lambda^2 - x^2 - y^2 - z^2}$, avec $\lambda = ct$.

Cet espace-temps est curieux. Il faut cependant nous y habituer.

Il est curieux pour deux raisons :

1° Il est à quatre dimensions. Les vulgarisateurs ont brodé à l'infini sur la quatrième dimension; des romans de science-fiction ont donné la méthode pour passer à travers cette dimension « supplémentaire ». Il ne faut cependant attacher aucun mystère à l'espace-temps : il est normal d'avoir trois dimensions relatives à l'espace et une supplémentaire représentant le temps (cela revient à dire tout simplement qu'il faut se donner quatre nombres pour localiser un événement dans l'espace et dans le temps);

2° La distance d'un point-événement à l'événement origine n'est pas

$$\lambda^2 + x^2 + y^2 + z^2,$$

comme une généralisation naturelle du théorème de Pythagore pourrait le faire penser, mais bien $\lambda^2 - x^2 - y^2 - z^2$. Ce fait est bien plus étonnant que le précédent. Si l'espace-temps était tel que l'on ait une généralisation naturelle du théorème de Pythagore, les théories relativistes n'existeraient pas. Il ne faut pas perdre de vue que l'hypothèse de la constance de la vitesse de la lumière et le principe de relativité ont imposé $s^2 = c^2t^2 - l^2$ et non pas $s^2 = c^2t^2 + l^2$.

En particulier, on a vu précédemment qu'il se peut que la distance spatio-temporelle de deux événements distincts représentés par deux points distincts de l'espace-temps soit nulle; c'est notamment le cas lorsque les deux événements sont le départ d'un signal en un endroit distant du premier de l. En effet $ct = l$, donc $s^2 = c^2t^2 - l^2 = 0$.

C'est en 1908 que le mathématicien russe Minkowski a imaginé l'éspace-temps. Citons une de ses phrases dans son ouvrage *Raum und Zeit (Espace et temps)* ; « A l'heure actuelle, l'espace et le temps considérés en eux-mêmes doivent disparaître comme des fantômes, et seule leur union peut posséder une individualité. »

Nous raisonnerons désormais le plus souvent possible en considérant l'espace-temps. L'espace-temps est le cadre approprié pour étudier la relativité.

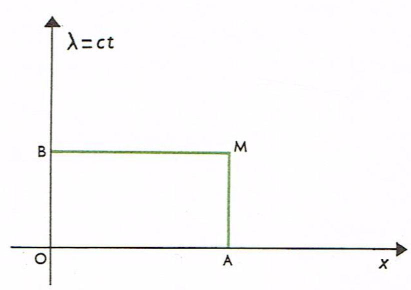
Nous avons là, il est vrai, une multiplicité quadridimensionnelle, mais nous sommes heureux d'avoir retrouvé un bon invariant. En mécanique classique, nous en avions deux : la distance spatiale et l'intervalle de temps; maintenant c'est plus simple, il n'y en a plus qu'un seul.

la vitesse dans l'espace-temps

Que l'espace-temps ait quatre dimensions n'est pas très grave; le seul ennui est l'impossibilité de le dessiner sur une feuille de papier (déjà pour représenter un objet à trois dimensions sur un tableau, on est obligé d'avoir recours à la perspective). Le fait essentiel est que dans cet espace-temps $s^2 = c^2t^2 - l^2$.

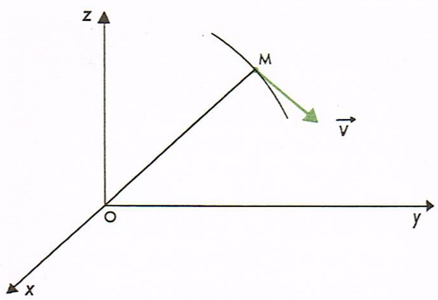
La transformation de Lorentz a été établie en considérant le mouvement d'un référentiel par rapport à un autre référentiel, et, dans l'exemple choisi, nous n'avons eu besoin que de deux dimensions : de la coordonnée x pour repérer le poste d'essence au bord de la route, et de la coordonnée t, temps indiqué par l'horloge du poste d'essence (puisque tout se passait sur l'axe Ox, les coordonnées y et z étaient toujours nulles). Si le repérage se faisait par rapport au référentiel « auto », x était remplacé par x', tandis que t l'était par t'; évidemment $c^2t^2 - x^2 = c^2t'^2 - x'^2$.

Pour simplifier, nous continuerons à raisonner sur un espace-temps à deux dimensions : une coordonnée spatiale et une coordonnée temporelle suffiront à repérer un événement. Les espaces seront portés en abscisses, et les valeurs, non pas de t, mais de $ct = \lambda$, en ordonnées. Nous avons là l'analogue d'un graphique de chemin de fer, avec cependant la modification consistant à substituer au temps t la grandeur ct. Les coordonnées de l'événement M sont OA $= x$ et OB $= ct$. Mais, attention! Prenons garde à ce schéma et surtout ne croyons



pas que $OM^2 = x^2 + c^2t^2$. Ici nous touchons bien la particularité de l'espace-temps : elle ne réside pas dans le nombre de ses dimensions, mais dans l'expression de la distance de deux points événements. Le graphique n'est qu'un schéma et non pas une figure de géométrie.

La géométrie de l'espace-temps n'est pas la géométrie habituelle (géométrie euclidienne), puisque le théorème de Pythagore ne s'applique pas. Cependant, c'est toujours de la géométrie plane



(les mathématiciens l'appellent « pseudo-euclidienne »).

En mécanique classique, le vecteur vitesse $\vec{v}$ d'un mobile M est la dérivée par rapport au temps du vecteur $\overrightarrow{OM}$. Si l'on se contente de repérer des mobiles sur une droite (autos sur une route rectiligne), le vecteur vitesse se réduit à la vitesse v telle que $v = \frac{x}{t}$ si le mouvement est un brave mouvement uniforme. Dans ce cas, une seule dimension de l'espace ordinaire est utilisée.

En relativité, nous imaginerons une définition de la vitesse faisant corps avec l'espace-temps. Cette nouvelle définition nécessitera de définir deux composantes : une u_1 relative à x, une autre u_0 relative à $\lambda = ct$ (dans l'espace-temps général, ce nouveau vecteur vitesse serait quadridimensionnel).

Comment définir u_1 et u_0? En mécanique classique, il n'y a pas d'ambiguïté : la vitesse est le quotient de l'espace parcouru par le temps t, le temps universel t. En relativité, cette dernière notion s'est évanouie; le temps varie d'un référentiel galiléen à un autre. Cependant, il reste associé à un événement M son temps propre τ, le temps mesuré par l'horloge liée à M. Souvenons-nous d'ailleurs que τ est invariant lors d'un changement de référentiel galiléen. Les relativistes peuvent alors définir les deux composantes u_1 et u_0 de la vitesse de l'événement dans l'espace-temps par $u_1 = \frac{x}{\tau}$ et $u_0 = \frac{\lambda}{\tau}$ (le mouvement de l'événement M est supposé uniforme pour éviter l'usage des dérivées).

A l'auto se déplaçant sur la route avec la vitesse v par rapport au référentiel « paysage » est liée la vitesse spatio-temporelle u de composantes u_1 et u_0. Evidemment, une liaison étroite existe entre u_1, u_0 et v. Elle est très simple à trouver. Il suffit de se souvenir que

$$\tau = t\sqrt{1 - \beta^2},$$

pour trouver

$$u_1 = \frac{x}{t} \cdot \frac{t}{\tau} = \frac{v}{\sqrt{1 - \beta^2}}$$

$$\text{et } u_0 = \frac{ct}{\tau} = \frac{c}{\sqrt{1 - \beta^2}} \qquad \left(\beta = \frac{v}{c}\right).$$

Ainsi, au nombre v sont associées les deux composantes de la vitesse relativiste u; si l'on envisage l'espace-temps dans toute sa généralité, on associe au trivecteur $\vec{v}$ (à trois composantes) un quadrivecteur spatio-temporel $\vec{u}$ (à quatre composantes : trois d'espace et une de temps). Les formules les liant sont analogues à celles qui ont été trouvées précédemment.

Une dernière question se pose à ce propos : quelle est la grandeur de la vitesse relativiste u? Oublions le théorème de Pythagore, qui risque de faire écrire $u^2 = (u_0)^2 + (u_1)^2$; ici, il nous faut appliquer la règle $s^2 = \lambda^2 - x^2$, et poser

$$u^2 = (u_0)^2 - (u_1)^2$$

$$= \frac{c^2 - v^2}{1 - \beta^2} = c^2 \frac{1 - \beta^2}{1 - \beta^2} = c^2.$$

La vitesse relativiste a donc une longueur constante c : quand la vitesse ($v = \beta c$) du mobile augmente, les composantes u_0 et u_1 diminuent, mais la différence de leurs carrés reste égale au carré de la vitesse de la lumière.

Ces considérations de cinématique relativiste sont un prélude à la dynamique relativiste, objet de l'exposé suivant.

$E = mc^2$, une formule explosive

Jusqu'ici nous n'avons parlé que de longueurs et d'intervalles de temps. Nous avons bâti l'espace-temps et la cinématique relativiste. Des vérifications expérimentales très intéressantes ont pu être effectuées : durée de vie des mésons, effet Doppler relativiste, expérience de Fizeau. Mais il faut bien avouer que, même si leur portée philosophique est grande, leur influence dans la vie courante est nulle. Il n'en sera pas de même des conséquences pratiques de cet exposé, puisque nous démontrerons l'existence de l'énergie nucléaire. Inutile d'insister; nous sommes tous au courant de ses méfaits et de ses bienfaits.

Abordons la dynamique relativiste, c'est-à-dire l'étude des forces provoquant les mouvements étudiés dans les chapitres précédents. La résistance qu'oppose la matière à toute modification de son mouvement se traduit par son inertie, sa masse. Masse, forces, énergie, tels sont les grands concepts dont nous nous préoccuperons dans ce passage passionnant à tous les points de vue.

masse et quantité de mouvement, que deviennent-elles ?

1° Rappelons ce qu'est la quantité de mouvement en mécanique classique. Un point matériel de masse m_0 animé de la vitesse $\vec{v}$ possède une quantité de mouvement $\vec{p} = m_0\vec{v}$.

La quantité de mouvement est donc un trivecteur spatial, comme la vitesse spatiale $\vec{v}$. En outre, m_0 est la masse d'inertie du mobile, et un postulat fondamental de la mécanique classique affirme que m_0 est une constante liée au mobile, indépendante de sa vitesse.

Va-t-il en être de même en relativité? Le résultat peut être prévu sans effectuer de calculs. La masse d'un corps traduit la difficulté qu'on a à le mettre en mouvement, ou à accélérer son mouvement s'il y en a un. Plus la masse est grande, plus il est difficile d'augmenter sa vitesse d'une quantité donnée.
En relativité, le fait qu'aucun corps ne puisse acquérir une vitesse supérieure à celle de la lumière nous amène à

conclure que sa résistance à une nouvelle accélération, c'est-à-dire **sa masse, doit croître d'une façon illimitée quand sa vitesse tend vers celle de la lumière.**

Donc, il est à prévoir que la masse m d'un corps en mouvement sera plus grande que sa masse au repos et qu'elle croîtra indéfiniment quand la vitesse v du corps tendra vers c. Il faut s'attendre à voir apparaître en dénominateur le fameux facteur $\sqrt{1 - \beta^2}$ qui tend vers zéro quand v tend vers c.

2° D'une façon plus précise, définissons la **quantité de mouvement relativiste** $\vec{\mathcal{P}}$. Elle sera liée naturellement à la quadrivitesse d'espace-temps $\vec{u}$ par la formule $\vec{\mathcal{P}} = m_0\vec{u}$. Si l'on se contente encore une fois de raisonner dans l'espace-temps simplifié à deux dimensions, le vecteur $\vec{\mathcal{P}}$ aura seulement deux composantes $\mathcal{P}_0$ et $\mathcal{P}_1$ ($\mathcal{P}_0 = m_0u_0$; $\mathcal{P}_1 = m_0u_1$).

Les formules établies à l'exposé précédent et donnant les valeurs u_0 et u_1 en fonction de la vitesse v du mobile permettent d'aboutir aux expressions suivantes :

$$\mathcal{P}_0 = \frac{m_0c}{\sqrt{1 - \beta^2}} \text{ et } \mathcal{P}_1 = \frac{m_0v}{\sqrt{1 - \beta^2}}.$$

En mécanique classique, l'expression de la quantité de mouvement d'une particule était $p = mv$. En mécanique relativiste, la composante spatiale $\mathcal{P}_1$ est égale à $\frac{m_0v}{1 - \beta^2}$.

Il est alors tout naturel de prendre l'expression $\frac{m_0}{\sqrt{1 - \beta^2}}$ comme la masse m de la particule animée de la vitesse v.

La formule $m = \frac{m_0}{\sqrt{1 - \beta^2}}$ secoue les conceptions traditionnelles des physiciens de 1906.

Evidemment, pour les phénomènes à l'échelle terrestre, cette formule n'a aucune importance. Même pour $v = 1\,000$ km/s, vitesse énorme pour nous, mais faible par rapport à c, on peut écrire $m = m_0$, en commettant une erreur inférieure à cinq millionièmes.

Mais, pour les particules élémentaires (électrons, protons, mésons, etc.) animées de vitesses de l'ordre de c, **la masse en mouvement sera très différente de la masse au repos.** Ainsi certains électrons existant dans les rayons cosmiques se meuvent à des vitesses égales à 99,98 p. 100 de la vitesse de la lumière; à de telles vitesses, leur masse est 1 000 fois celle de la masse au repos.

Les vérifications expérimentales de la formule $m = \frac{m_0}{\sqrt{1 - \beta^2}}$ ne se comptent plus à l'heure actuelle; elles reposent soit sur l'étude du mouvement d'une particule chargée dans un champ électrique ou dans un champ magnétique, soit sur l'étude du choc élastique de deux particules.

Comme il a été vu par ailleurs, la mesure de $\frac{e}{m}$, le rapport de la charge d'un électron à sa masse, se déduit de la déviation de cet électron par un champ électrique ou par un champ magnétique. Il est inutile de revenir ici sur la description du phénomène et des appareils utilisés. Les résultats expérimentaux prouvent que m varie avec la vitesse de l'électron suivant la formule relativiste. Il suffit de citer à l'appui les expériences de Kaufmann en 1901, de Bucherer en 1908, de Guye et Lavanchy en 1915, de Nacken en 1935. D'ailleurs, sans anticiper sur la suite de l'ouvrage, il faut signaler que les cyclotrons, accélérateurs de particules lourdes, fonctionnent mal ou ne fonctionnent plus quand ces particules atteignent des vitesses élevées. Il faut les remplacer par des synchrotrons, qui, eux tiennent compte de la variation de la masse avec la vitesse.

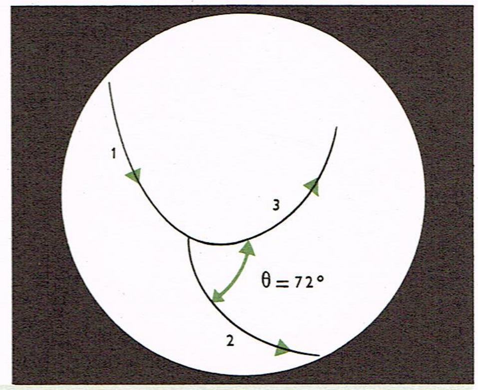
Le choc élastique de deux particules P_1 et P_2 est lui aussi très instructif. Une particule de masse m animée d'une certaine vitesse v bouscule une autre particule au repos; si le choc est « élastique », la quantité de mouvement totale de P_1 et de P_2 reste inchangée avant et après le choc, de même leur énergie totale. Si l'on se place dans le cadre de la mécanique classique, il est facile de démontrer que l'angle θ est droit. Si l'on admet la doctrine relativiste (la quantité de mouvement est alors égale à $p = \frac{m_0v}{\sqrt{1 - \beta^2}}$), on démontre que l'angle θ est obligatoirement aigu.

L'expérience vérifie magnifiquement cette prévision; il suffit pour s'en convaincre d'admirer le cliché obtenu par M. Leprince-Ringuet, qui représente le choc relativiste électron-électron, l'électron incident ayant une vitesse de 279 000 km/s; l'angle θ est de 72°. De telles expériences mettent en évidence la validité de ces théories. Mais ce n'est pas tout : les principes extrêmement féconds de la relativité restreinte ont servi de base pour édifier de nombreuses autres théories, telle celle de l'électron magnétique de Dirac, qui entraîne *ipso facto* la prévision du spin de l'électron. **Toute la physique des particules est résolument et fondamentalement relativiste.**

enfin l'équation explosive !

L'équation fondamentale de la mécanique relativiste diffère de celle de la mécanique classique : le temps absolu a été supprimé purement et simplement, le théorème sur l'addition des vitesses a été profondément remanié, l'accélération d'un mobile n'est donc plus la même pour deux repères galiléens distincts. En outre, nous venons de découvrir que la masse n'était plus un invariant fondamental. La formule





Au point A, l'électron relativiste 1 heurte de plein fouet un électron tranquille du gaz et le projette avec une grande vitesse. Après le choc (analogue à celui de billes de billard), les deux électrons partent selon 2 et 3, et l'angle de leurs trajectoires est inférieur à l'angle droit, qui serait de rigueur en mécanique classique.

Doc. Leprince-Ringuet.

classique $\vec{F} = m_0\vec{\gamma}$ est à rejeter si l'on adopte une attitude relativiste. Par quelle équation la remplacer ? Pour y parvenir, le relativiste s'adresse à l'espace-temps, cadre bien approprié au développement de ses théories. Pour aboutir à ses fins, notre homme n'emploiera pas des méthodes tellement révolutionnaires. Il se laissera guider par l'intuition et il essaiera tout simplement d'écrire une formule quadridimensionnelle en accord avec la géométrie de l'espace-temps. Ce qui est sûr, c'est que, pour des vitesses faibles par rapport à celle de la lumière, il devra retrouver les formules classiques. Pour suivre le cheminement de sa pensée, précisons bien les bases de la mécanique classique.

1° L'équation vectorielle $\vec{F} = m_0\vec{\gamma}$ équivaut à trois équations scalaires (c'est-à-dire à des relations entre des grandeurs exprimées par des nombres), comme disent les mathématiciens. Il est préférable d'écrire cette formule en faisant intervenir la quantité de mouvement $\vec{p} = m_0\vec{v}$ du mobile. En se souvenant que $\vec{\gamma} = \frac{d\vec{v}}{dt}$, on peut donner à la relation précédente la forme $\vec{F} = \frac{d\vec{p}}{dt}$.

2° Jusqu'ici, seule la notion de force est intervenue. C'est une notion aussi vieille que les hommes ; mais une force n'a d'intérêt pratique que si elle déplace le corps auquel elle est appliquée, que si elle travaille. La force $\vec{F}$ agissant sur le corps considéré augmente progressivement sa vitesse ; le travail T de la force est égal à l'augmentation de l'énergie cinétique $\frac{1}{2}m_0v^2$ du mobile. On peut alors écrire pour l'augmentation d'énergie :

$$\Delta T = \vec{F}.\vec{l},$$

$\vec{l}$ étant le déplacement du point d'application de la force $\vec{F}$ supposée constante. (ΔT désigne le produit scalaire des vecteurs $\vec{F}$ et $\vec{l}$, c'est-à-dire le nombre $Fl \sin \theta$.)

Ce théorème ne traduit pas autre chose que la conservation de l'énergie mécanique. A quelle sauce ce théorème va-t-il être mangé par les relativistes ? Les relativistes nous ont habitués à la destruction de concepts fondamentaux, mais, pour eux comme pour les physiciens classiques, la conservation de l'énergie est un dogme. Ils écriront eux aussi, nous le verrons, $\Delta T = \vec{F}.\vec{l}$. Oui, mais au fur et à mesure que la vitesse du mobile augmente, sa masse augmente. Il est alors clair que T ne peut plus être égal à

$$\frac{1}{2}m_0v^2.$$

C'est là que les relativistes vont se séparer des mécaniciens classiques.

Sachant que la mécanique classique se résume en deux formules, une formule vectorielle (à trois composantes) où intervient la force et une formule scalaire où intervient la notion d'énergie, que vont faire les relativistes ? Oh ! un raisonnement bien simple. Nous avons une formule vectorielle plus une formule scalaire, soit quatre équations ; l'espace-temps est là, avec ses quatre dimensions. Inventons une formule unique, mais quadridimensionnelle, vraie quel que soit le référentiel galiléen adopté. Posons

$$\vec{\mathcal{F}} = \frac{d\vec{\mathcal{P}}}{d\tau}.$$

$\vec{\mathcal{F}}$ sera le quadrivecteur force, $\vec{\mathcal{P}}$ le quadrivecteur quantité de mouvement, et τ le temps propre déjà utilisé pour définir le quadrivecteur vitesse $\vec{u}$. Evidemment, les trois composantes spatiales de $\vec{\mathcal{F}}$ doivent tendre pour des vitesses faibles vers les trois composantes de la force classique ; de même pour $\vec{\mathcal{P}}$. Il est à prévoir que la quatrième composante temporelle de $\vec{\mathcal{F}}$ conduira au théorème de l'énergie cinétique remanié par les relativistes.

Il faut effectuer maintenant les calculs. Ils ne présentent aucune difficulté, surtout si, pour simplifier, on se contente du schéma d'espace-temps à deux dimensions. Voici les résultats.

Les trois composantes spatiales de

$$\vec{\mathcal{F}} = \frac{d\vec{\mathcal{P}}}{dt} \text{ donnent } \vec{\mathcal{F}} = \frac{d}{dt}\left(\frac{m_0\vec{v}}{\sqrt{1-\beta^2}}\right),$$

et la quatrième composante

$$\Delta T' = \Delta\left(\frac{m_0c^2}{\sqrt{1-\beta^2}}\right) = \Delta(mc^2) = \vec{F}.\vec{l}.$$

La première formule est très satisfaisante. Si t est le temps mesuré dans le référentiel « laboratoire » par rapport auquel le mobile a la vitesse $\vec{v}$, $\vec{\mathcal{F}}$ reste la dérivée par rapport au temps t de la quantité de mouvement mv, mais avec $m = \frac{m_0}{\sqrt{1-\beta^2}}$, conformément aux résultats précédents, bien vérifiés expérimentalement. **La seconde formule mérite un examen particulièrement approfondi.** Elle va se révéler explosive...

la matière, immense réservoir d'énergie

Ainsi l'application de la force $\vec{F}$ modifie la vitesse — donc la masse — de la particule, de telle manière que la variation $\Delta T' = \Delta(mc^2)$ soit égale au produit scalaire des deux vecteurs $\vec{F}$ et $\vec{l}$ (déplacement du point d'application de $\vec{F}$). Par analogie avec la mécanique classique, on est tenté d'écrire que mc^2 est l'expression relativiste de l'énergie cinétique du mobile.

Mais, attention ! T' n'est défini que par ses variations $\Delta T'$. Il en résulte que l'énergie cinétique n'est connue qu'à une constante près. De même, un alpiniste, mesurant à sa peine la dénivellation qu'il a effectuée, ne peut en déduire l'altitude qu'il vient d'atteindre que s'il connaît celle dont il est parti. Il nous faut donc écrire

$$T' = mc^2 + K = \frac{m_0c^2}{\sqrt{1-\beta^2}} + K$$

(constante encore indéterminée) et nous préoccuper de la détermination de cette constante.

Est-il possible de l'évaluer, et comment ?

L'alpiniste peut toujours définir une « altitude zéro » ayant un sens physique (niveau de la mer, défini par le marégraphe de Marseille, ou, mieux, centre de la Terre). Il est clair que, si nous voulons que T' soit une énergie de vitesse, T' doit être nulle lorsque la vitesse l'est aussi. Imaginons donc un repère galiléen entraîné par le mobile même. Par rapport à lui, $T' = 0$, $\beta = 0$, d'où il résulte que K, loin d'être nulle, vaut $-m_0c^2$. La formule que nous venons de démontrer nécessite bien d'être mise en caractère gras :

$$\mathbf{T' = mc^2 - m_0c^2.}$$

C'est la formule explosive annoncée. Malgré la modestie de son apparence, elle est d'une portée immense et couronne l'édifice de la relativité restreinte. Ses conséquences pratiques ? Toutes les conséquences de l'existence de l'énergie nucléaire.

Avant de voir tout cela, réfléchissons encore à notre équation, qui peut être mise sous la forme $T' = W - W_0$.

$W = mc^2$ est l'énergie du mobile animé de la vitesse v, tandis que $W_0 = m_0c^2$ est l'énergie du mobile au repos. $T' = W - W_0$ exprime le principe de la conservation de l'énergie en termes relativistes.

$W_0 = m_0c^2$ est l'énergie que possède

un corps de masse m_0 posé à terre, sur le sol, au repos. Ne confondons pas cette énergie avec l'énergie potentielle d'un lac de haute montagne. L'eau du lac, si l'on établit une conduite forcée, tombera dans la vallée et fera tourner des turbines. Mais cette eau, une fois dans la vallée, ne peut plus fournir d'énergie notable; les mécaniciens classiques disent qu'elle a perdu son énergie potentielle. $W_0 = m_0c^2$ mesure l'énergie que recèle toute portion de matière de masse m_0. C'est l'énergie d'une goutte d'eau dans une soucoupe. Si l'homme trouve le moyen d'annihiler, d'anéantir, de désintégrer cette goutte d'eau, il récupérera l'énergie m_0c^2, et cette quantité est colossale.

Prenons un kilogramme d'eau et calculons l'énergie E en puissance dans cette faible quantité.

$$c = 3.10^8\ m/s;$$

$E = 1 \times 9 \times 10^{16} = 9.10^{16}$ joules, ou encore 9.10^{13} kilojoules, c'est-à-dire 90 000 000 000 000 de kilojoules, soit environ une année de consommation d'énergie en France. Malheureusement (ou heureusement), l'homme ne sait encore désintégrer qu'une très faible fraction de matière donnée, dans des réactions dites « nucléaires ». Néanmoins, ce faible pourcentage permet de libérer des quantités d'énergie importantes utilisables pratiquement (usines nucléaires de l'E. D. F., sous-marins et navires à propulsion atomique, etc.).

En mécanique classique, la quantité de mouvement d'une particule est donnée par la formule $\vec{p} = m_0\vec{v}$. En relativité, il faut envisager le quadrivecteur quantité de mouvement $\vec{\mathcal{P}}$: ses trois composantes spatiales forment le trivecteur $\vec{\mathcal{P}} = m\vec{v}$, tandis que la composante temporelle $\mathcal{P}_0 = mc$ peut s'écrire maintenant $\mathcal{P}_0 = \frac{E}{c_0}$, puisque $E = mc^2$.

Les physiciens parlent du quadrivecteur impulsion-énergie d'une particule. Par rapport à un repère galiléen donné, ses composantes sont mv_X, mv_Y, mv_Z et $\frac{E}{c}$.

Evidemment, si le référentiel galiléen change, les composantes du vecteur $\vec{\mathcal{P}}$ sont modifiées suivant une transformation de Lorentz.

Si tout corps au repos de masse m_0 est équivalent à l'énergie m_0c^2, réciproquement toute énergie W est équivalente à une masse « morte » m_0 telle que

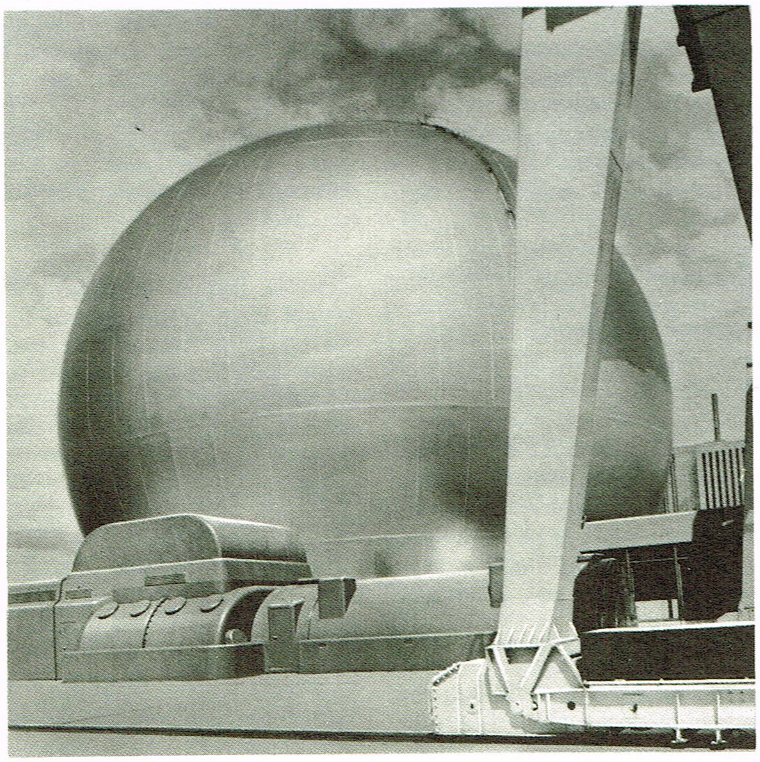
$$m_0 = \frac{W}{c^2}.$$

Ainsi, si la désintégration de la matière libère de l'énergie, à l'opposé, de l'énergie peut se matérialiser; et les physiciens ont observé ce phénomène.

un instant de réflexion

Nous sommes impatients de prendre contact avec toutes ces merveilleuses conséquences de la relativité restreinte, ce que nous ferons aux articles ultérieurs. Néanmoins, en arrivant au terme de ce chapitre, il n'est pas inutile de faire un retour en arrière et d'essayer de porter un jugement critique sur la théorie.

A la base de la relativité, il y a une critique serrée des notions classiques d'espace et de temps, critique rendue nécessaire par les résultats négatifs de l'expérience de Michelson. Et si l'expérience de Michelson avait été mal faite, mal interprétée? Faire reposer une théorie si révolutionnaire sur des bases expérimentales aussi fragiles semblait une gageure. Sacrifier des notions fondamentales pour une expérience, quelle témérité!... Les antirelativistes de 1920 ne se sont pas fait faute de brandir de tels arguments. Les relativistes ont répliqué : notre théorie serait sans doute née même si Michelson n'avait pas tenté son expérience, tellement le désir d'unifier les lois de la nature est grand chez les physiciens. Au XIXe siècle, deux disciplines scientifiques se sont disputé la première place : la mécanique classique et l'électromagnétisme. Les lois de la mécanique classique étaient invariantes par rapport à une transformation de Galilée, tandis que les équations de Maxwell, clef de voûte de l'électromagnétisme, étaient invariantes par rapport à une transformation découverte par Lorentz que nous connaissons bien maintenant. Cet antagonisme ne pouvait durer. L'une des deux théories devait céder devant l'autre. Pendant un moment, des physiciens comme Hertz ont torturé l'électromagnétisme pour l'adapter à la mécanique, mais sans grand succès. Les lois de l'électromagnétisme l'ont finalement emporté, aidées, il faut le dire, par le résultat négatif de l'expérience de Michelson; et le principe fondamental de la dynamique a été modifié pour devenir invariant dans toute transformation de Lorentz. L'essor de la relativité ne s'est jamais ralenti et chaque jour de nouvelles preuves expérimentales viennent appuyer la théorie. Il suffit de se reporter à l'exposé de physique nucléaire donné dans cet ouvrage pour en être convaincu.



Réacteurs nucléaires de l'E. D. F., à Chinon.

Phot. Ehrmann.

vers la relativité générale...

La relativité restreinte n'a considéré que des systèmes particuliers : les systèmes galiléens. Elle affirme que dans tout système galiléen la lumière se propage en ligne droite avec la même vitesse dans toutes les directions, que dans chaque système on peut accorder les horloges par l'usage d'un signal optique, que cette vitesse est une constante universelle, que les formules de transformation de coordonnées sont celles de Lorentz, enfin que toutes les lois physiques restent les mêmes quand on change de repère galiléen.

la relativité restreinte est insuffisante

Que dire alors des phénomènes se produisant par rapport à des repères non galiléens, par rapport à des repères accélérés. L'espace-temps que nous avons bâti dans les chapitres précédents sera-t-il encore adéquat? Nous savons que si nous sommes sur un disque tournant, une nouvelle force apparaît du fait de la rotation du disque : la force centrifuge. Cette force existe réellement, et il semble bien que la rotation du disque soit un phénomène absolu. Et voilà un beau pavé jeté dans la mare de la relativité!

La relativité restreinte, d'autre part, a ignoré complètement un phénomène pourtant important : la gravitation. Sur la Terre, tous les corps abandonnés à eux-mêmes tombent suivant la verticale du lieu où ils se trouvent (la pesanteur est un cas particulier de la gravitation universelle). La Lune tourne autour de la Terre parce qu'elle est attirée par elle. Les planètes, sous l'action d'un champ de gravitation, tournent autour du Soleil. L'attraction universelle se manifeste aussi par le mouvement des étoiles doubles. **Ignorer la gravitation est vraiment impardonnable pour une théorie.**

Einstein, dès 1905, a eu conscience de ce défaut et, pendant de nombreuses années, il a réfléchi au problème. Il a fait la remarque simple suivante : partout où existe un champ de pesanteur, un champ d'accélération existe. A partir de là il a bâti une théorie de la gravitation révolutionnaire et a montré qu'un univers réel avec des astres doués de masses, donc **un univers avec gravitation, ne peut obéir aux lois de la géométrie euclidienne** (est nécessairement courbe).

La relativité restreinte apparaît alors comme une théorie limite, valable rigoureusement dans un univers sans masses (si tant est qu'un univers vide ait un sens). Heureusement, loin des masses importantes, et pour tout phénomène où la gravitation joue un rôle négligeable, la relativité restreinte constitue une excellente approximation. C'est le cas, très important, du mouvement des particules élémentaires (électrons, protons, mésons, etc.) dans les champs électrique et magnétique. Toute la mécanique des grands accélérateurs de particules (cyclotrons, bêtatrons, synchrocyclotrons, etc.) peut donc s'appuyer sur les seuls principes de la relativité restreinte.

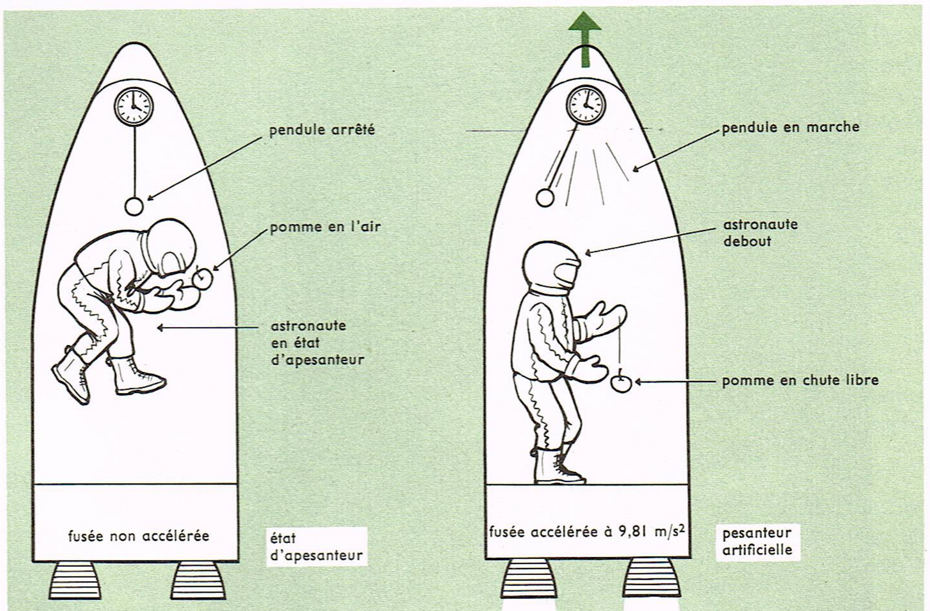
Einstein a publié ses travaux sur la relativité générale en 1916. Nous n'en donnerons qu'une idée, car il est impossible de pénétrer dans le détail de la théorie. Nous pensons que les calculs associés aux théories de la relativité restreinte sont élémentaires, mais l'étude de la relativité générale fait appel à des êtres mathématiques compliqués, les **tenseurs,** que nous ne pouvons aborder dans cet ouvrage.

Sachons, pour terminer cette introduction à la relativité générale, que des expériences de plus en plus nombreuses sont imaginées par les physiciens pour vérifier la théorie, et que les résultats sont positifs. Nous évoquerons ces expériences à la fin de ce chapitre.

le principe d'équivalence

Nous connaissons tous l'expérience du tube de Newton. Dans un tube, où l'on a fait préalablement le vide, on abandonne à eux-mêmes, au sommet du tube, une bille d'acier, une plume, un morceau de liège, par exemple. Ils tombent tous les trois également vite et arrivent au bas du tube au même instant.

Tous les corps, qu'ils soient lourds ou légers, tombent dans le vide avec la



apesanteur et pesanteur artificielles **(v. page suivante)**

même vitesse si les conditions sont les mêmes. L'accélération est indépendante de la force qui s'exerce sur les corps.

Ce fait familier est extraordinaire. C'est une propriété extrêmement remarquable du champ de pesanteur (ou de gravitation), qui n'appartient pas au champ électrique et magnétique. Dans de tels champs, des charges différentes prennent des accélérations différentes.

Donc nous conclurons que **le champ de pesanteur se manifeste par un champ d'accélération.**

Imaginons maintenant (p. 175) une fusée interplanétaire en mouvement uniforme dans l'espace, assez loin de toute planète pour supposer qu'il n'y ait dans son enceinte aucune force de gravité. Tous les objets situés à l'intérieur d'une telle fusée, et les cosmonautes avec eux, seront sans poids et flotteront librement dans la fusée (on négligera les attractions qu'exercent les éléments de la fusée les uns sur les autres).

A un certain moment, les cosmonautes mettent les réacteurs en marche. La fusée commence à se déplacer en gagnant peu à peu de la vitesse. Son mouvement s'accélère, et supposons que cette accélération soit de 9,81 m/s^2. Que va-t-il se passer à l'intérieur de la cabine ? Tous les objets vont par inertie tendre à se déplacer vers le plancher. Le cosmonaute sera à l'aise, debout, les pieds bien d'aplomb. S'il laisse tomber la « pomme de Newton », elle prendra dans la cabine un mouvement uniformément accéléré, comme sur la Terre. Si en O on suspend un pendule de 1 m de longueur, il battra la seconde comme à Paris. Les cosmonautes seront comme sur la Terre. Ils seront incapables de trouver la moindre différence entre les phénomènes qui se passent dans leur cabine accélérée et les phénomènes de pesanteur sur la Terre.

En somme, il y a équivalence entre un système accéléré sans gravitation (fusée) **et un système galiléen avec gravitation** (la Terre en première approximation).

Ce principe a été énoncé pour la première fois par Einstein lui-même.

La relativité restreinte étudie les lois physiques à bord de référentiels gali-

Au sommet d'une trajectoire parabolique, les occupants d'un avion se trouvent, pendant quelques secondes, en état d'apesanteur.



Phot. U. S. I. S.

léens sans gravitation. La relativité générale se préoccupe des systèmes galiléens avec gravitation, ou, ce qui revient au même, des systèmes accélérés. Ainsi, quel que soit le référentiel utilisé, les lois physiques s'expriment de la même manière à condition de tenir compte de la gravitation. Les accélérations, comme l'espace, comme le temps, ont perdu leur caractère absolu, puisque un système accéléré peut être remplacé par un système sans accélération, mais avec gravitation. Le référentiel « auto » freine... Le passager est précipité contre le pare-brise... « Tiens ! pourra-t-il dire, le champ de gravitation terrestre est devenu horizontal, je tombe sur le pare-brise. » Revenons à notre passager du Boeing 707 se prélassant à 10 000 m d'altitude. Brutalement, l'avion amorce une montée accélérée en chandelle. Quelle est la réaction du voyageur, qui, souvenons-nous, a été laissé endormi à Orly? Il peut très bien dire : « Je suis écrasé sur mon fauteuil, je pèse donc plus lourd, la pesanteur terrestre a augmenté. »

méthodes, ancienne et moderne, pour supprimer la pesanteur

Peut-on supprimer la pesanteur à la surface de la Terre? Bien sûr! Il suffit de trouver le système de référence accéléré adéquat et de s'y « mettre ».

La méthode ancienne consiste tout bonnement à s'enfermer dans un ascenseur tombant en chute libre. On flotte dans la cage, mais une telle descente ne peut qu'être aussi dangereuse qu'éphémère. Un tel état est d'ailleurs créé dans un avion piquant avec une accélération de 9,81 m/s².

La méthode moderne est de devenir cosmonaute dans un satellite artificiel. Quand le satellite est mis en orbite, il constitue un système accéléré qui tourne autour de la Terre. La force centrifuge équilibre exactement le poids. Le cosmonaute est dans un état d'apesanteur. Cet état d'apesanteur a joué un tour à John Glenn, le premier cosmonaute américain qui ait fait trois fois le tour de la Terre : il a emmêlé les fils qui tenaient son appareil photographique, n'a pu le sortir et a enragé de ne pas pouvoir prendre les vues sensationnelles de la Terre à haute altitude. (Mais on peut, en faisant tourner la fusée sur elle-même, y créer un champ d'accélération, que le cosmonaute ne peut distinguer d'un champ de pesanteur auquel il est habitué.)

Puisqu'il y a équivalence entre un champ de pesanteur et un système accéléré, il suffit d'étudier les phénomènes physiques dans un système accéléré : les résultats se transposeront au cas gravitationnel. Comme la géométrie utilisable dans les systèmes accélérés va se révéler non euclidienne, il nous faut réfléchir quelques instants aux « espaces courbes ».

les espaces courbes

Nous savons tous ce qu'est une ligne courbe, une surface courbe, mais il est difficile de se représenter ce que peut être un espace courbe à trois dimensions. Et pourquoi ? En y réfléchissant un peu, on se rend compte qu'une ligne est courbe, qu'une surface est courbe parce que nous ne nous trouvons pas dessus. Nous les voyons de l'extérieur. Nous pouvons regarder une orange, une selle de cheval, mais, en ce qui concerne l'espace à trois dimensions, comment nous rendre compte qu'il est courbe puisque nous sommes dedans!

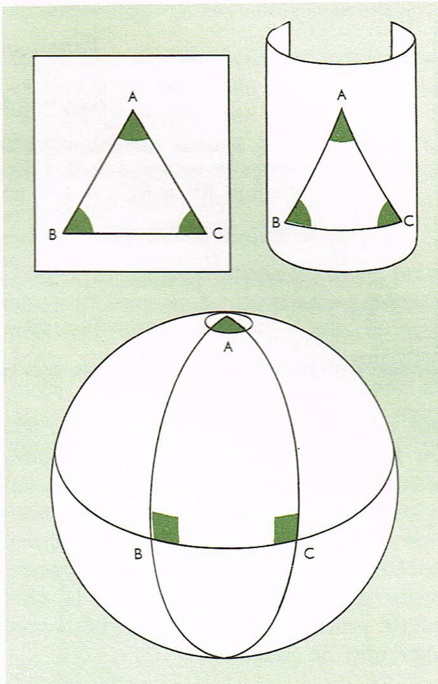
Cette notion de courbure est intuitive, et la notion précise, mathématique, de la courbure est un peu différente : une surface est courbe si les propriétés des figures géométriques tracées sur elle sont différentes de celles qui sont tracées sur un plan. **La courbure, c'est l'écart aux règles de la géométrie d'Euclide.**

Traçons, par exemple, un triangle sur une feuille de papier; la somme de ses trois angles fait 180°. Enroulons la feuille de papier pour en faire un cyclindre, un cône, ou toute autre forme plus compliquée. La somme des trois angles du triangle fera toujours 180°.

La géométrie de la surface n'a pas changé du fait de ces déformations. Les surfaces rendues courbes au sens ordinaire du mot seront aussi « plates » qu'un plan, pour le mathématicien.

Il n'en est plus de même pour une sphère ou pour une selle de cheval. On ne peut appliquer sur elles sans la déchirer une feuille de papier. Pour y arriver, il faudrait une feuille de caoutchouc. Ces surfaces sont vraiment courbes.

Essayons de dessiner un triangle sur une sphère. Une première question se pose : quelle courbe tracée sur la sphère sera l'équivalente de la droite du plan? Rappelons la propriété fondamentale de la droite : c'est le plus court chemin pour aller d'un point à un autre d'un plan (on dit que la droite est une géodésique de l'espace euclidien).



La somme des trois angles du triangle dessiné sur un papier est de 180°, que la feuille soit plate ou arrondie. Mathématiquement parlant, la surface du papier enroulé n'est pas courbe. En revanche — écart à l' « euclidiérité » —, la somme des angles du triangle A B C formé par deux méridiens et l'équateur est supérieure à deux angles droits.

Quelles courbes sur la sphère jouissent de la même propriété? Ce sont les grands cercles (intersections de la sphère par les plans passant au centre de la sphère). Le plus court chemin de Paris à New York est le grand cercle passant par Paris et New York. Donc un triangle sphérique est constitué par trois grands cercles qui se coupent. Considérons un de ces triangles constitué par l' « équateur » et deux « méridiens » de la Terre (prise comme exemple de sphère). En B et C les angles sont droits (les méridiens coupent l'équateur à angle droit). Donc, obligatoirement,

$$A + B + C > 2 \text{ droits.}$$

C'est une propriété générale des triangles sphériques : la somme de leurs angles est supérieure à deux droits.

Si l'on étudiait la somme des angles d'un triangle dessiné sur une selle de cheval, on montrerait qu'elle est inférieure à deux droits.

La géométrie de la surface d'une sphère, de la surface d'une selle n'est pas euclidienne. La surface d'une sphère, la sur-

Équation d'Einstein.

$$R_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} R = k.T_{\alpha\beta}.$$

$R_{\alpha\beta}$, $g_{\alpha\beta}$, $T_{\alpha\beta}$ sont les composantes d'êtres appelés « tenseurs ». Les indices α et β peuvent prendre les valeurs 1, 2, 3, 4. Ces tenseurs étant symétriques (ce qui veut dire par exemple que $R^{12} = R^{21}$), il y a pour chaque tenseur 10 composantes. Il y a 10 équations d'Einstein.

$R_{\alpha\beta}$ est le tenseur de courbure de l'espace-temps, R la courbure scalaire. $R_{\alpha\beta}$ et R traduisent donc la courbure de l'espace-temps.

Les dix $g_{\alpha\beta}$ sont la généralisation du potentiel de gravité de la physique classique. $T_{\alpha\beta}$ est le tenseur de matière. Il traduit les propriétés de la matière (et du champ électromagnétique au point de l'espace-temps considéré).

face d'une selle de cheval sont courbes. De même, si notre espace est courbe, certaines mesures permettront sans nul doute de mettre en évidence cette courbure. Evidemment, ce seront des mesures extrêmement fines, car, depuis de nombreux siècles, les théorèmes de la géométrie euclidienne sont vérifiés avec beaucoup de précision.

Après cette incursion dans le domaine des mathématiques, nous allons revenir à la physique et constater avec étonnement que la géométrie dans un système de référence accéléré n'est pas euclidienne.

le rotor de la mort : un système accéléré

Aux fêtes foraines, une baraque s'installe parfois, sur laquelle on lit : « Le rotor de la mort. Venez éprouver des sensations fortes. » (V. phot. p. 35.)

Le rotor de la mort est un cylindre qui peut tourner de plus en plus vite autour de son axe. Les amateurs de sensations fortes se tiennent au milieu du fond du cylindre, celui-ci étant au repos. Le rotor se met en marche. Il s'agit de tenir le plus longtemps possible au centre, mais rapidement, la vitesse croissant, les amateurs, soumis à la force centrifuge, sont plaqués le long de la paroi verticale du cylindre. Le plancher s'abaisse alors et seule la force centrifuge empêche les expérimentateurs de tomber.

Revenons aux choses sérieuses et considérons des expérimentateurs vivant sur un disque plat en rotation uniforme autour d'un axe passant par son centre. Le disque est assez grand et le coefficient de frottement des expérimentateurs sur le disque est suffisant pour qu'ils puissent s'y arrêter en un point quelconque. (Remarquons que nous sommes dans un cas similaire puisque la Terre tourne autour de son axe. La rotation est suffisamment lente pour que le phénomène soit négligeable, mais il n'en existe pas moins.) Supposons, en outre, qu'il n'y ait aucun repère extérieur permettant à ces personnes de s'apercevoir que le disque tourne. Evidemment, elles ressentent toutes les effets de la force centrifuge.

Quelle va être la géométrie adoptée par ces habitants d'un système accéléré (nous nous souvenons que tout mouvement de rotation, même uniforme, est un mouvement accéléré)?

Prenons l'exemple suivant : un « habitant du disque » désire mesurer le rapport de la longueur de la circonférence C à son rayon OA. Si le disque était arrêté, il trouverait évidemment 2π.

Le disque est en mouvement par rapport au sol. Pour un observateur lié au référentiel « sol », le rapport de la longueur de la circonférence à son rayon est bien 2π, comme d'ailleurs pour l'habitant du disque lorsque ce dernier est arrêté.

Un géomètre du disque mesure le rayon OA avec un mètre. Il trouve le même nombre que l'observateur fixe. En revanche, lorsqu'il va mesurer un morceau de circonférence, sa mesure sera en désaccord avec celle du Terrien fixe, puisque ce morceau de circonférence est animé d'une vitesse v par rapport à la Terre. le Terrien trouvant 2π comme rapport de la longueur de la circonférence au rayon, l'habitant du disque en rotation ne peut pas trouver 2π.

un dialogue de sourds

Imaginons d'ailleurs un dialogue entre les deux hommes.

LE TERRIEN. — Sur votre disque, la géométrie euclidienne n'est pas valable, votre référentiel est mal choisi; prenez donc mon bon référentiel « Terre », et la géométrie d'Euclide sera sauvée. Votre disque a un mouvement absolu; c'est un système accéléré; les lois de la physique y sont compliquées à étudier.

L'HOMME DU DISQUE. — Mon disque ne tourne pas; c'est vous qui tournez par rapport à moi. Mon référentiel est aussi bon que le vôtre. Euclide est une vieille barbe!

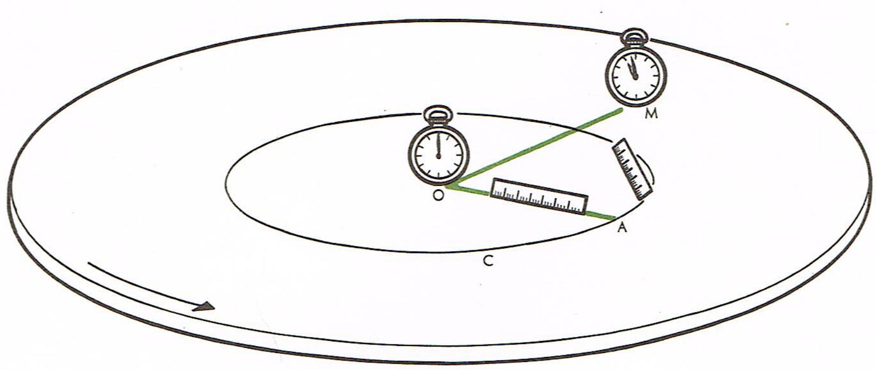
LE TERRIEN. — Pourtant, vous sentez une force étrange vous pousser vers l'extérieur. C'est la preuve absolue que votre disque tourne.

L'HOMME DU DISQUE. — Et puis après? Cette force étrange, c'est un champ de gravitation qui agit sur mon disque, déforme mes chaînes d'arpenteur et modifie le rythme de mes horloges...

La discussion menace de s'éterniser. Les deux interlocuteurs ont aussi raison l'un que l'autre. Leur querelle est sans objet.

La géométrie du disque n'est pas euclidienne. L'espace pour les habitants du disque est courbe.

Autre phénomène : soit deux montres mises à l'heure en O. L'une est amenée en M. Elle aura une certaine vitesse



par rapport au point O, donc retardera sur celle qui est restée au centre.

Il existe cependant une force attirant tous les corps du disque vers l'extérieur : la force centrifuge. Est-ce que les habitants vont attribuer l'existence de cette force à la rotation du disque ? Pas obligatoirement. Ils ne savent pas qu'ils tournent. Aussi peuvent-ils considérer cette force comme une force de pesanteur. Nous, nous constatons une attraction vers le centre de la Terre; eux constatent une attraction vers la périphérie du disque.

Quelles sont alors les conclusions générales que l'on peut déjà tirer de ces exemples particuliers ?

Comme il y a équivalence entre un système accéléré sans gravitation et un système avec gravitation, toute région de l'espace où il existe des forces de gravitation est courbe et la courbure de l'espace est d'autant plus grande qu'il y a à proximité des masses importantes.

La géométrie naturelle à adopter est une géométrie non euclidienne. Construite par Riemann vers 1860, elle est dite « riemannienne ». (Les mathématiciens ne se préoccupent pas de la réalité physique. Ils édifient des constructions logiques et en tirent des conséquences appelées « théorèmes ». Mais rien ne laisse prévoir que des mathématiques extrêmement abstraites ne puissent servir un jour au physicien.)

Avec Einstein, nous admettrons qu'**un champ gravitationnel n'est que la manifestation physique de la courbure de l'espace.**

Toute cette théorie de la relativité générale a été mise sous forme mathématique. C'est un magnifique édifice très logique, très beau par sa clarté et sa cohérence interne.

la sanction de l'expérience : de la courbure des rayons lumineux à l'effet Mössbauer

Au voisinage de masses importantes, dans un champ de gravitation intense où la courbure de l'espace est notable, la lumière ne se propage plus en ligne droite (puisqu'elle décrit des géodésiques, qui, dans un espace courbe, ne sont plus des trajectoires rectilignes).

Pour établir simplement ce phénomène, nous ne prendrons pas une région de l'espace où existe un champ de gravitation, mais un système de référence accéléré, puisque nous savons qu'il y a équivalence entre les deux.

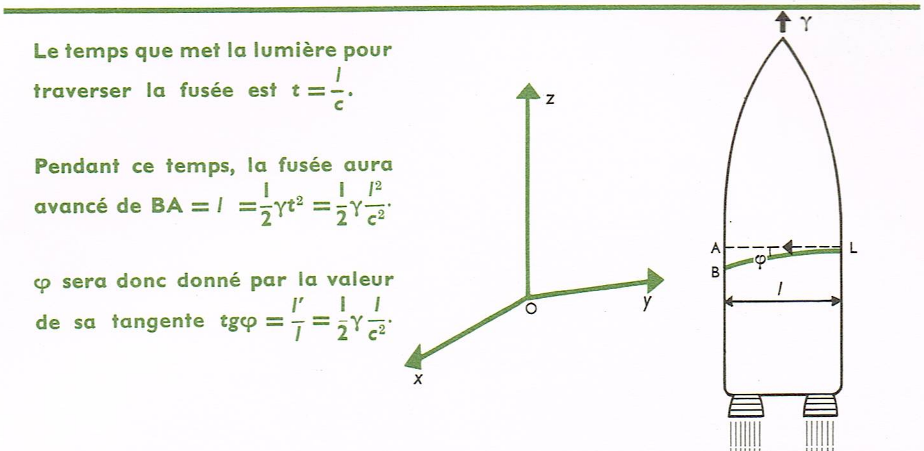
Revenons à la fusée interplanétaire animée d'un mouvement uniformément accéléré, et étudions comment un rayon lumineux se propage dans la fusée.

Si la fusée est au repos ou en mouvement rectiligne uniforme par rapport à un référentiel galiléen O*xyz*, la lumière se propage en ligne droite. Le rayon lumineux issu horizontalement de la lampe L frappe un écran en A.

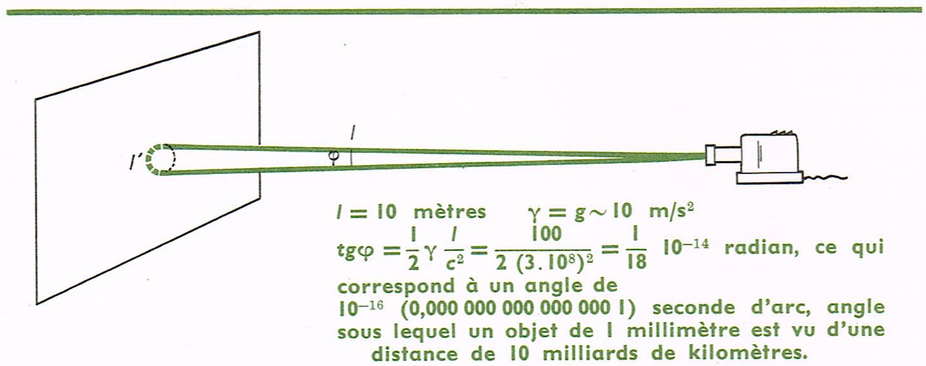
Si la fusée est en mouvement accéléré par rapport à O*xyz*, la lumière se courbe et vient frapper l'écran en B. Pour un observateur situé en O, cette trajectoire est le résultat de la combinaison du mouvement rectiligne et uniforme de la lumière à la vitesse c et du mouvement accéléré de la fusée, et l'on peut calculer l'angle φ.

Remarquons que l'on aboutirait à la même conclusion à partir d'un raisonnement différent. Un rayon lumineux transporte de l'énergie, donc a une masse inerte m telle que $m = \frac{W^2}{c^2}$. Or, d'après le principe de l'équivalence, masse inerte et masse pesante sont identiques, et la lumière possédant une masse pesante est déviée par un champ de gravitation. Un tel raisonnement est au fond le même que le précédent, mais le principe d'équivalence est présenté différemment.

Dans l'exemple précédent, ni la masse de la Terre ni la longueur parcourue n'étaient suffisantes pour que la courbure soit observable. Adressons-nous alors aux rayons lumineux venant des étoiles et rasant le Soleil. Le calcul montre que la déviation atteint 1,74'' (1).



I. Déviation de la lumière due au champ d'accélération de la fusée.



II. Déviation d'un rayon lumineux due au champ de pesanteur terrestre.

S'il n'y a aucun astre entre une étoile E_1 et la Terre T, la lumière chemine en ligne droite; mais si sur le trajet s'interpose le Soleil, le rayon lumineux suit la trajectoire courbe E_1T et un observateur terrestre verra l'étoile dans la direction E'_1 au lieu de E_1. Et l'angle

(1) Une seconde d'arc se note par le signe '' (1° 2' 52'' exprime un angle de 1 degré 2 minutes et 52 secondes), alors que « s » désigne une seconde de temps (1 h 2 mn 52 s se lit 1 heure 2 minutes 52 secondes). Ces deux notations sont trop souvent confondues, notamment dans les comptes rendus sportifs.

de déviation mesure 1,74″. Le déplacement de l'étoile E_1 sur la sphère céleste est donc tout à fait mesurable. Malheureusement, les étoiles au voisinage du Soleil sont invisibles, et il faut attendre de trop rares éclipses de Soleil, pour atténuer son éclat. Les astronomes ont effectué des mesures de déviation lors des éclipses de 1919, 1923, 1928, et ont trouvé 1,75″, 1,72″, 1,82″. Les résultats sont en excellent accord avec la théorie de la relativité. Notons au passage la difficulté extrême des mesures. 2″ d'arc, ce n'est pas beaucoup (1″ d'arc est l'angle sous lequel on voit 1 cm à 20 km de distance!). Admirons la finesse et la précision des mesures astronomiques.

Sans pénétrer dans l'édifice mathématique de la relativité générale, nous voudrions faire, malgré tout, quelques remarques. Pour la relativité générale comme pour la relativité restreinte, le cadre naturel de la théorie est l'espace-temps. Mais, cette fois-ci, il faut envisager un espace-temps courbe. L'espace-temps de la relativité restreinte était « plat ». Les droites étaient le plus court chemin d'un point à un autre. En relativité générale, **dans l'espace-temps courbe, les trajectoires ne sont plus rectilignes, ce sont des géodésiques.** Et, évidemment, ce sont des lignes courbes.

D'ailleurs, tout corps libre décrit lui aussi dans l'espace-temps courbe une géodésique. Ce qui nous amène à signaler une autre vérification de la théorie.

Les applications de la loi de Newton et de la mécanique classique ont obtenu un succès quasi universel. Toute la mécanique céleste en témoigne. Cependant, vers 1850, les astronomes mirent en évidence un léger désaccord dans le mouvement de la planète Mercure (planète gravitant tout près du Soleil). Ils trouvèrent une anomalie dans l'avance du périhélie (point de la trajectoire elliptique le plus rapproché du Soleil) de Mercure de 42″ par siècle. C'était vraiment peu, et, cependant, cet écart chagrinait beaucoup les savants; ils avaient beau accumuler les hypothèses les plus artificielles, apportant même des termes correctifs à la loi de la gravitation universelle, inventant même de nouvelles planètes, aucune interprétation ne pouvait être trouvée à ce désaccord, petit certes, mais certain.

La théorie de la relativité générale apporte l'explication du désaccord. Au lieu de dire « le Soleil exerce une certaine force qui agit directement sur les planètes et leur fait décrire des ellipses autour de lui », il vaut mieux dire « la masse du Soleil courbe l'espace-temps E autour de lui, et les trajectoires spatio-temporelles des planètes sont des géodésiques tracées dans E ». En appliquant cette idée et en faisant les calculs correspondants, on trouve effectivement une avance de 42″ par siècle du périhélie de Mercure, conformément aux résultats expérimentaux. Tout naturellement la relativité générale justifie une anomalie qui ne peut être expliquée d'une manière simple par la théorie classique de Newton.

●

Nous savons que la variation du champ de gravitation va influencer la marche d'une horloge, puisque nous savons qu'il en est de même pour un système accéléré. Pour faire un calcul simple, reprenons l'exemple du disque qui tourne avec une vitesse angulaire.

Le calcul montre qu'il y a bien un ralentissement dans le rapport $1 - \frac{W}{c^2}$: plus le potentiel de gravitation est grand, plus l'horloge est ralentie.

Si nous avons deux horloges identiques, l'une sur la Terre, l'autre sur le Soleil, elles auront des marches différentes. De même, tout atome vibrant sera ralenti : il émettra donc une radiation plus « rouge ».

On a effectivement observé ce déplacement vers le rouge dans les spectres du Soleil et de quelques autres étoiles (en particulier si ces étoiles sont très denses), et les mesures réalisées sont en bon accord avec les prévisions de la théorie.

Serait-il possible de mettre en évidence le ralentissement du temps dû à la gravitation à la surface de la Terre? Jusqu'en 1960 on n'y croyait pas, mais un nouveau phénomène découvert par le physicien Mössbauer en 1958 a permis de réaliser des expériences positives.

De l'osmium activé dans des circonstances très particulières peut émettre un rayonnement γ dont les raies, d'une extrême finesse (effet Mössbauer), peuvent être déterminées avec une précision telle qu'un phénomène aussi petit que celui qu'on se propose de détecter peut être mis en évidence sans être noyé par l'effet Doppler-Fizeau d'agitation.

Effectivement, aux laboratoires de Princeton et de Harwell (U. S. A.), ainsi qu'en U. R. S. S., les physiciens — utilisant des techniques particulièrement raffinées — ont pu constater que le simple fait d'élever l'appareil de 20 mètres (ce qui entraîne une infime variation du champ de gravitation terrestre) suffit à entraîner un effet mesurable.

Cette série d'expériences comptera dans l'histoire de la relativité. En effet, c'est la première fois que sur Terre, dans un laboratoire, une vérification de la relativité générale a été tentée avec succès. La déviation de la lumière par le Soleil, l'avance du périhélie de Mercure, le décalage vers le rouge des raies solaires sont des observations astronomiques, imposées par la nature à l'homme. Les expériences à propos de l'effet Mössbauer ont été conçues et imaginées par les physiciens. C'est un immense progrès; peut-être même un jour pourra-t-on satelliser une horloge et mesurer l'effet correspondant.

encore quelques problèmes à résoudre

Aujourd'hui la relativité est admise par tous les physiciens sans exception aucune. Il n'en a pas toujours été ainsi, et certains se souviennent des joutes homériques entre Langevin et Painlevé, le premier partisan enthousiaste de la relativité, le second adversaire acharné de cette théorie.

La simplicité de la relativité restreinte, la cohérence même de la relativité générale sont, avec les expériences multiples réalisées, les appuis solides de la relativité.

La relativité générale a, en particulier, fait de la théorie de la gravitation une véritable théorie du champ, et en cela elle n'est que la digne fille spirituelle de Maxwell, de Hertz et des physiciens du XIXe siècle. Mais, comme pour toute œuvre humaine, son élaboration est loin d'être terminée, et de nombreux problèmes restent posés. Les mathématiciens et les physiciens actuels orientent notamment leurs recherches vers les points suivants :

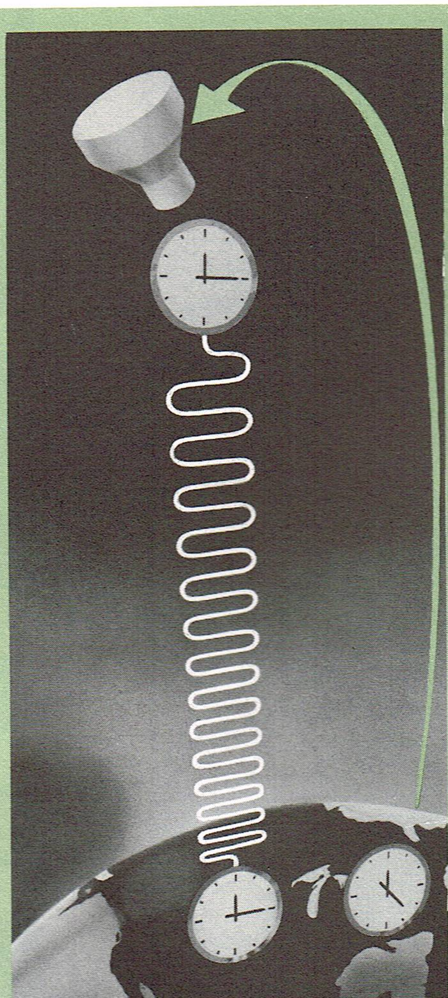
1° On sait maintenant que les phénomènes de gravitation se propagent à la même vitesse que la lumière. L'étude théorique des ondes gravitationnelles

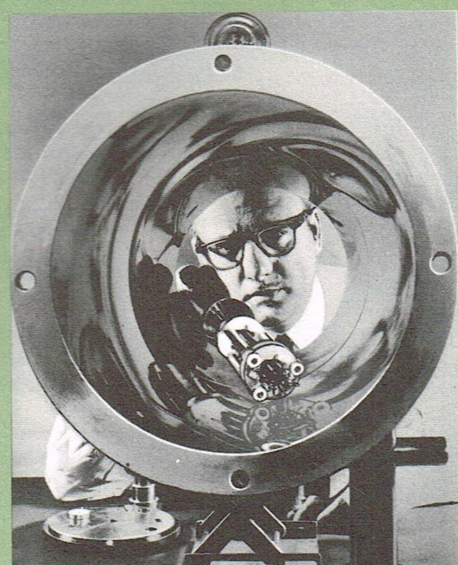
est bien avancée, mais il reste à les mettre en évidence expérimentalement. Des dispositifs sont à l'étude, mais rien de net n'a encore été publié;

2° A côté du champ gravitationnel existe le champ électromagnétique. Ce dernier a servi de guide pour l'établissement de la relativité restreinte, puisque la propagation de la lumière en est une conséquence. En relativité générale, il ne joue pas le même rôle que le champ gravitationnel. Des tentatives pour fondre le champ électromagnétique et le champ de gravitation en un seul champ ont été faites. Des résultats importants ont été obtenus, mais bien des ombres subsistent;

3° La liaison de la relativité générale avec la mécanique ondulatoire est délicate à faire. Actuellement, la question est étudiée en France par M. Lichnerowicz et son école;

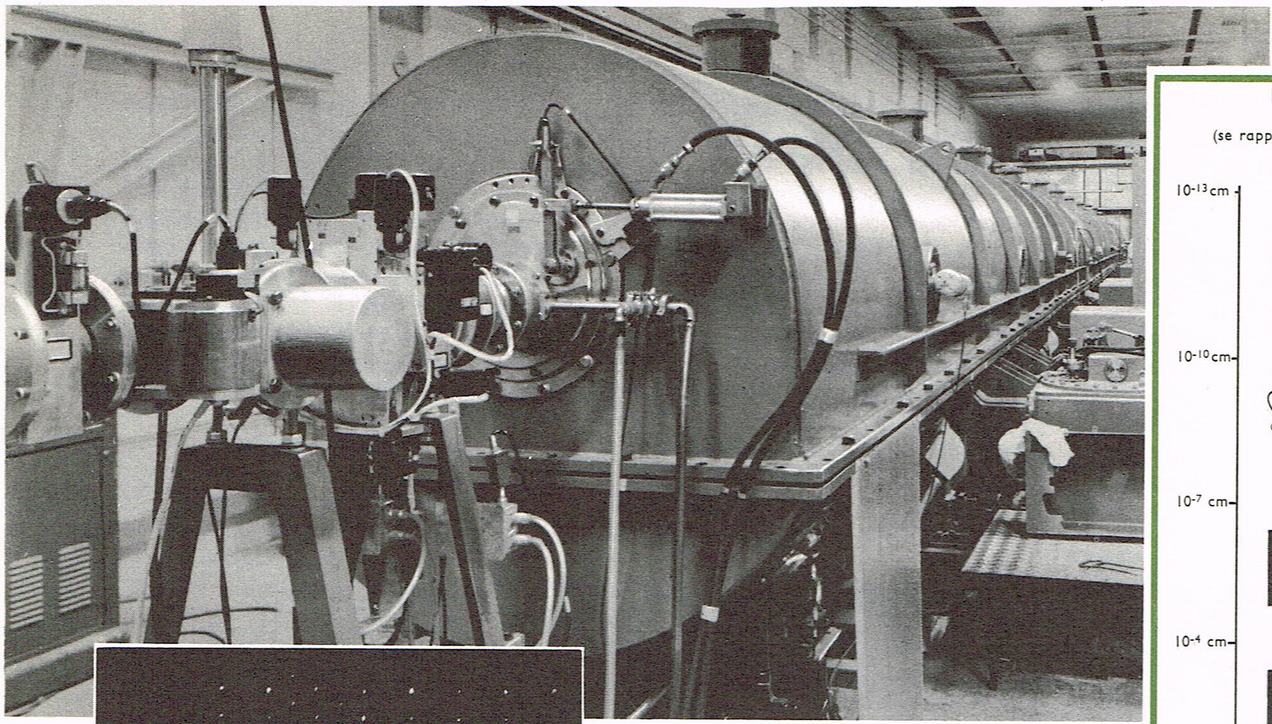
4° Enfin, un des prolongements les plus passionnants de la relativité générale est l'étude de la forme de l'univers. **L'univers est-il infini? Se referme-t-il sur lui-même?** De nombreuses hypothèses cosmologiques ont été avancées. Elles sont toutes osées, mais n'est-ce pas le propre de l'homme de toujours se surpasser, de vouloir toujours savoir plus, pour mieux comprendre l'univers dont il est une partie infime, mais capable de pensée?

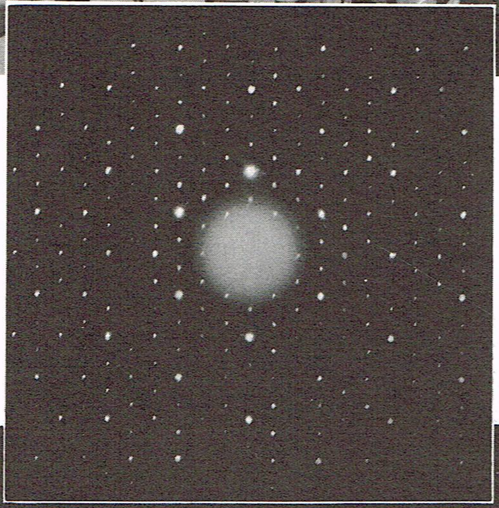




L'inversion de la position de l'atome d'azote dans une molécule d'ammoniac se fait à une fréquence parfaitement connue 23870,127 MHz indépendante des variations de pression et de température des ébranlements, voire du vieillissement de la matière. D'où la mise au point d'horloges incroyablement précises. D'où aussi l'idée de placer l'une de ces horloges sur un satellite (en parfait synchronisme avec une horloge identique restée au sol) pour vérifier expérimentalement que le temps s'écoule différemment dans les deux cas, conformément aux théories de la relativité.

Phot. U. S. I. S.



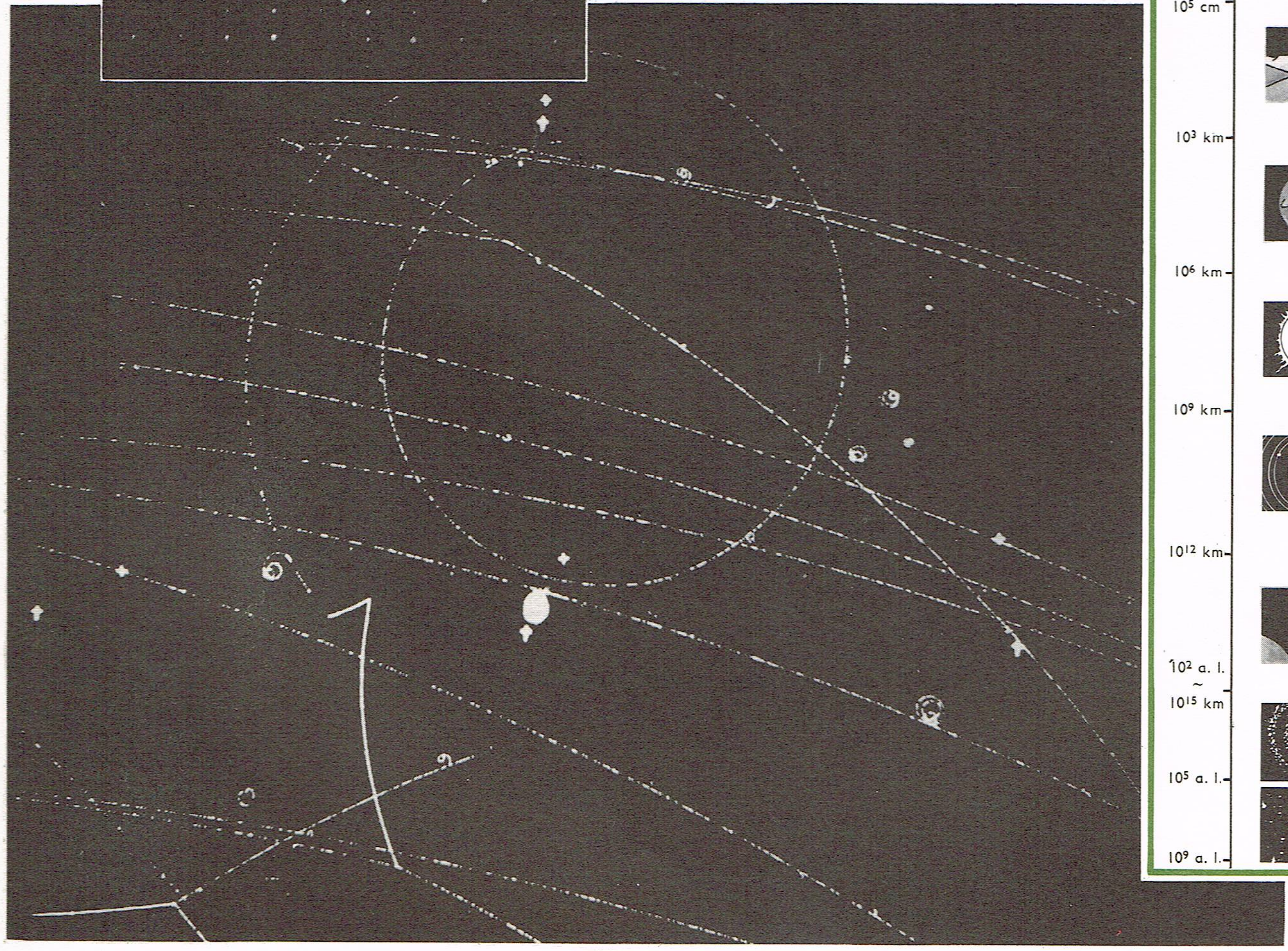


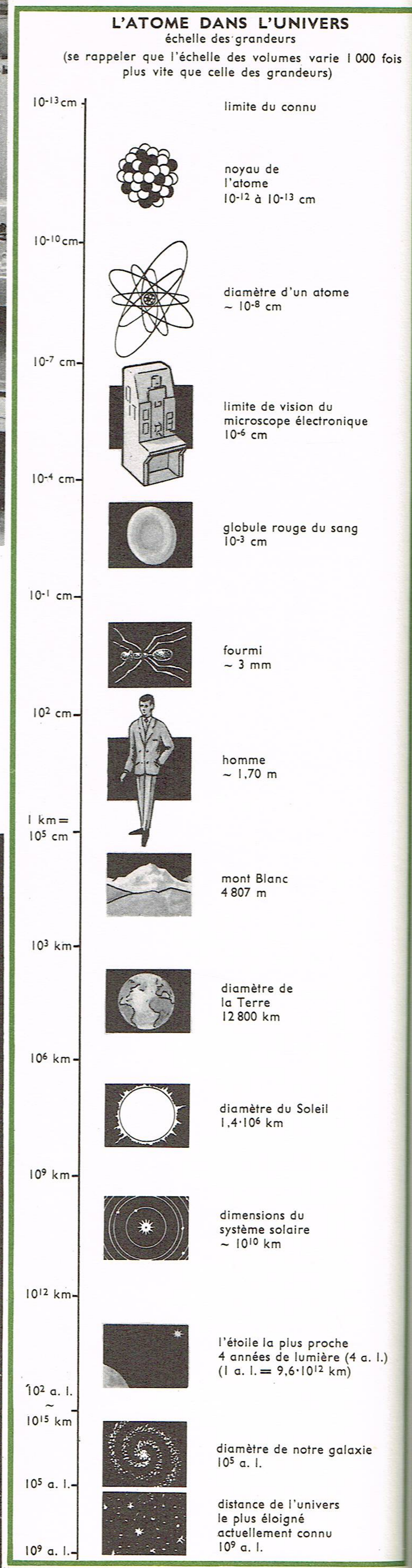
Ci-dessus, **cet accélérateur linéaire fournit le faisceau des protons qui seront injectés dans le synchrotron du CERN;**

Ci-contre, **électrons diffractés par une grille cristalline ;**

Ci-dessous, **trajectoires diverses de particules dans une chambre à bulles ;**

En page de droite, **désintégration nucléaire d'un noyau d'argent par les rayons cosmiques.**

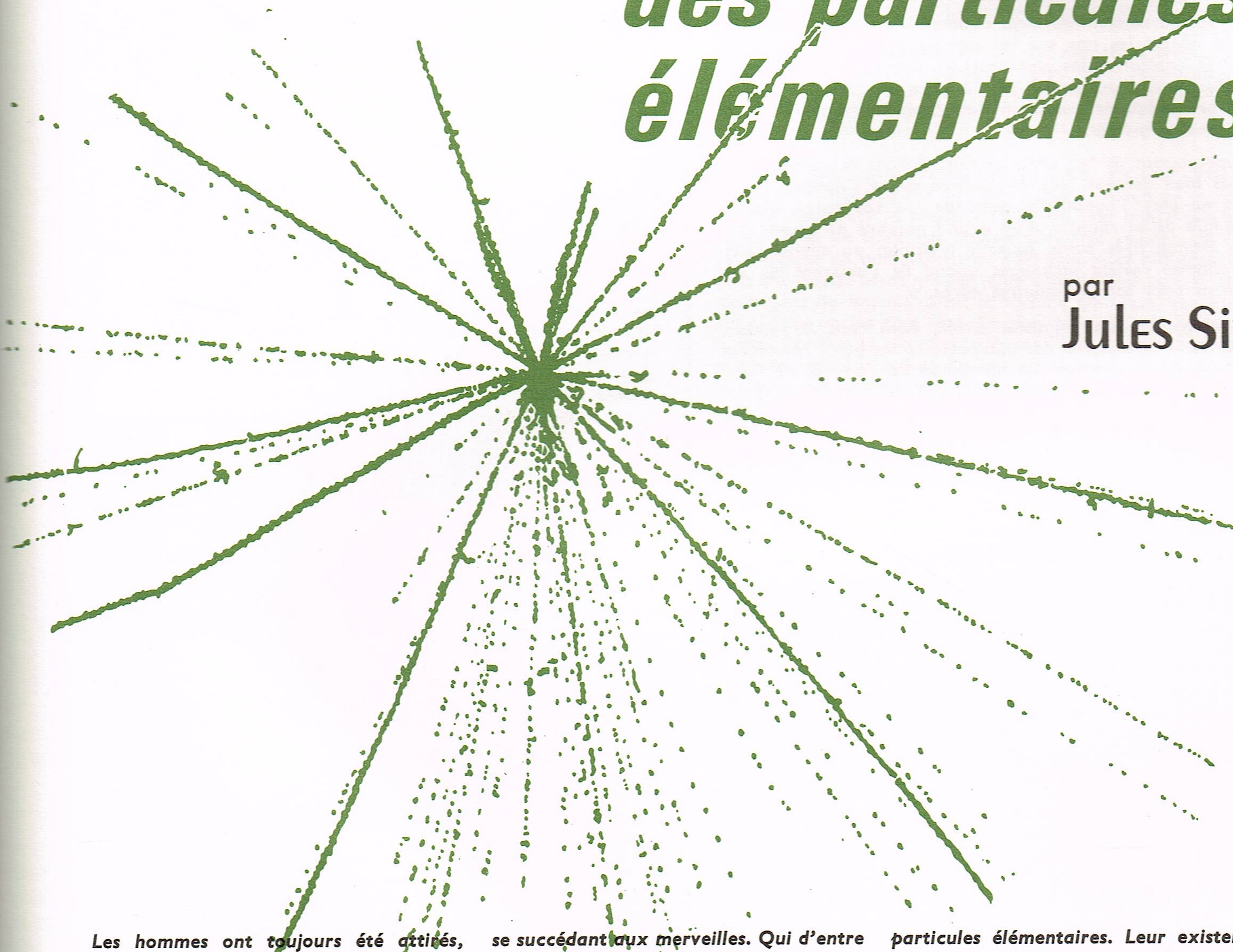




LE MONDE AGITÉ

des particules élémentaires

par
Jules Six



Les hommes ont toujours été attirés, sinon fascinés, par le monde de l'infiniment petit. Ils ont toujours espéré y trouver la clé à nombre de questions concernant l'univers. Les philosophes de l'Antiquité échafaudèrent des théories de la constitution de la matière qui voulaient rendre compréhensible le monde plein de mystères qui s'offrait à leurs yeux. Mais c'est au XVIIe siècle, avec l'invention du microscope, que les hommes commencèrent à entrer dans les secrets de l'infiniment petit. A vrai dire, l'attente n'a pas été déçue, les merveilles se succédant aux merveilles. Qui d'entre nous n'a pas admiré une simple goutte de sang dans l'oculaire d'un microscope? Et aujourd'hui le microscope électronique permet de grossir jusqu'à 160 000 fois, nous faisant voir jusqu'aux virus les plus petits et nous amenant peut-être même aux frontières de la vie.

C'est l'une des merveilles des sciences atomique et nucléaire que d'avoir non seulement permis la construction même du microscope électronique, mais encore de nous avoir fait découvrir — sans les voir — les atomes, les noyaux et les particules élémentaires. Leur existence à tous est sûre, et pourtant l'atome est 100 fois plus petit que le plus petit virus, le noyau 100 000 fois plus petit que l'atome lui-même. Nous sommes même entrés dans les secrets des lois qui règlent tout ce petit monde, règles paradoxales, sinon choquantes pour notre bon sens. C'est l'histoire de ces découvertes successives que nous nous proposons de retracer ici, depuis celle des atomes et des électrons, à la fin du XIXe siècle, jusqu'à celles toutes récentes qui ont fait reculer les frontières de l'inconnu.

atomes et électrons

premières découvertes : l'atome de Rutherford

les atomes des premiers chimistes

Préoccupés de classer les corps chimiques en catégories et de les faire réagir les uns sur les autres, quelquefois de façon brutale et explosive, les chimistes du XVIIIe et du début du XIXe siècle arrivèrent petit à petit à la conclusion qu'il existait un nombre réduit de corps simples, à partir desquels étaient formés tous les autres.

Les combinaisons de tels corps se faisant dans des proportions de poids bien définies, une explication claire s'imposa à eux de plus en plus, l'hypothèse atomique : « Les corps simples sont formés de la réunion d'un grand nombre de grains très fins ou atomes (très petits, puisque invisibles au microscope), et ces atomes restent inchangés quand ils s'unissent les uns aux autres pour former d'autres corps. »

Le grand travail des chimistes a été, dans cette hypothèse, d'établir les « formules » des composés et les poids atomiques relatifs des corps simples. La formation de l'eau et de l'eau oxygénée à partir de l'hydrogène H et de l'oxygène O s'explique ainsi aisément en écrivant :

molécule d'eau :
OH_2 = 1 atome d'oxygène + 2 atomes d'hydrogène,
molécule d'eau oxygénée :
O_2H_2 = 2 atomes d'oxygène + 2 atomes d'hydrogène,
et l'oxygène excédentaire de l'eau oxygénée vient expliquer de façon simple ses propriétés désinfectantes.

De telles formules servirent aussi à déterminer les poids atomiques relatifs des différents corps simples, qui se révélèrent être à peu près des multiples entiers de celui de l'hydrogène, le corps le plus léger entre tous. L'hypothèse atomique venait ainsi ordonner la diversité de la nature. Mendéléev, vers 1871, partant d'une classification des corps simples par poids atomiques croissants, trouva qu'on pouvait les répartir par périodes et grouper ainsi des corps aux propriétés chimiques semblables. Il prédit même l'existence de corps simples alors inconnus, qui furent effectivement découverts plus tard.

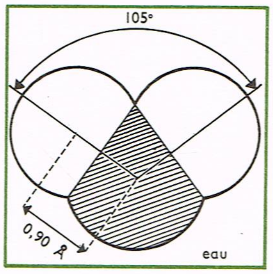
La réalité de l'atome s'imposa aux chimistes alors qu'ils ne savaient pratiquement rien de ses dimensions, sinon qu'il était très petit. On devait commencer à mieux le connaître à la suite de la découverte des ions et des électrons.

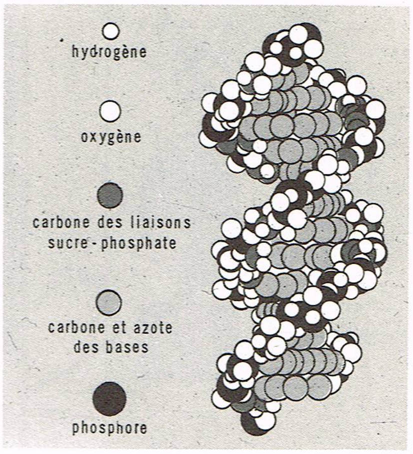
ions et électrons

Les électrons ont été découverts en 1897 par J. J. Thomson, dans un appareil qui est un peu l'ancêtre de nos modernes tubes de télévision.

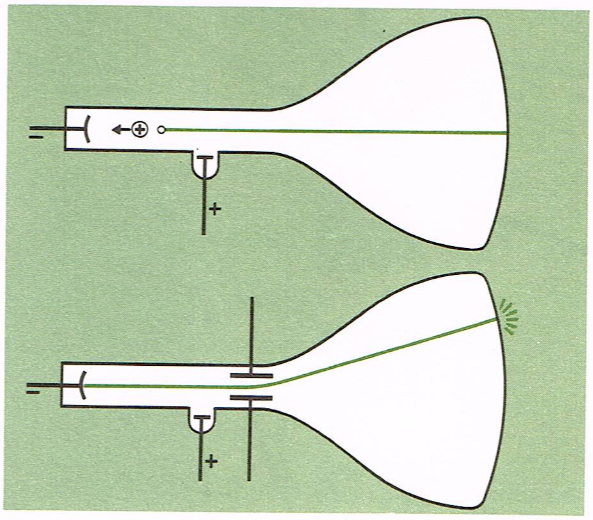
Un tube de verre contient un gaz raréfié. Sous l'influence d'une tension électrique, les atomes se scindent en deux parties : des « ions » de charge électrique positive (+ e), qui se dirigent vers la cathode (électrode négative), et des ions de charge électrique négative et opposée (— e), qui, manquant l'anode située dans un renflement latéral, se dirigent vers l'écran fluorescent et l'illuminent. Ces ions peuvent être déviés par des champs électriques et magnétiques (principe du balayage en télévision). Plus exactement, l'action combinée d'un champ

Diversité des édifices moléculaires.

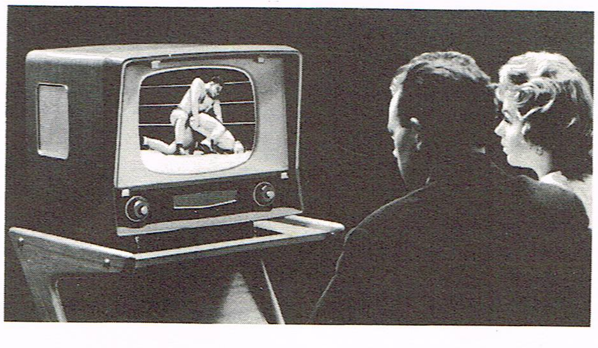




La molécule de l'eau et celle de l'A. D. N. (acide désoxyribonucléique).



Dans un tube de télévision, le courant d'électrons est de l'ordre du microampère (un millionième de coulomb par seconde) et correspond à un nombre d'électrons frappant par seconde l'écran de $\frac{10^{-6}}{1,6.10^{-19}} \sim 10^{13}$ (soit dix millions de millions).



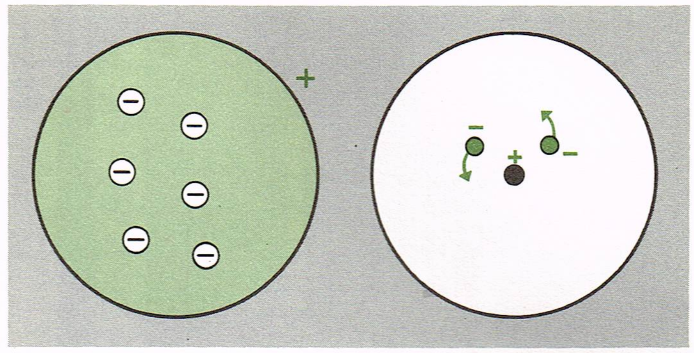
magnétique et d'un champ électrique donne une déviation qui permet de mesurer la vitesse et le rapport $\frac{e}{m}\left(\frac{\text{charge électrique}}{\text{masse}}\right)$ de l'ion.

C'est là, décrit schématiquement, le principe de ce qu'on appelle un **spectrographe de masse :** si la charge e de tous les ions est la même, on peut mesurer leurs masses relatives (en mesurant $\frac{e}{m}$). J. J. Thomson s'est ainsi aperçu que les ions négatifs étaient environ 2 000 fois plus légers que le plus léger des ions positifs H^+ (ion d'hydrogène). **Ainsi fut découvert l'électron.**

Une expérience très ingénieuse et très difficile (expérience célèbre de Millikan en 1909) permit ensuite de mesurer directement la charge électrique de l'électron, qui est extrêmement faible : $e = 1{,}6.10^{-19}$ coulomb. On peut alors mesurer la masse de l'électron et des différents atomes, puis leur taille. Qu'il suffise de dire ici qu'un gramme d'hydrogène contient le nombre fantastique de 6.10^{23} atomes d'hydrogène!

découverte du noyau atomique : Rutherford (1911)

En dépit de son nom, depuis la découverte de l'électron l'atome ne peut plus être considéré comme indivisible. Comment sont alors distribuées dans l'atome les charges positive et négative est une question qui a passionné les chercheurs. J. J. Thomson croyait que l'atome était une pastèque modèle réduit, mais Rutherford trouva que la charge positive de l'atome devait, au contraire, être concentrée dans un très petit volume. Il en acquit la certitude en faisant des expériences sur le rayonnement α de la radio-activité naturelle d'atomes lourds découvert par Becquerel. Un élément radio-actif émet un rayonnement α, c'est-à-dire des atomes d'hélium ionisés (atomes d'hélium sans électrons). Une fente faite dans un écran de plomb laisse



Dans l'atome de Thomson, les électrons sont dans une matière chargée positivement de façon uniforme.

Le modèle d'atome suggéré par Rutherford : les électrons gravitent autour du noyau positif très petit.

passer un faisceau mince de ces rayons α, qui sont révélés par un écran fluorescent sur une petite plage, image de la fente. Si l'on interpose sur le trajet du faisceau une mince feuille d'or (il est facile de faire des feuilles d'or très minces, de l'ordre du centième de millimètre), les particules sont diffusées parfois avec des angles assez grands (elles sont déviées par les charges électriques du noyau). Dans le modèle atomique de Thomson, ces grands angles sont inexplicables, puisque les charges électriques positive et négative voisinent trop. Au contraire, dans le modèle de Rutherford les rayons ne sont pas gênés par les électrons très légers (pas plus qu'une boule de billard n'est gênée par les grains de poussière) et peuvent s'approcher assez près du noyau et subir une déviation importante.

De plus, ces rayons ne sont pas tous déviés, et la fréquence des déviations donne des renseignements sur les dimensions du noyau, qui se révèlent être 100 000 fois plus petites que celles de l'atome.

CLASSIFICATION DE MENDÉLÉIEV DES ÉLÉMENTS REMANIÉE ET MODERNISÉE

Chaque corps simple est caractérisé par deux nombres :

1. Son numéro atomique, qui indique simplement sa position et qui représente en même temps le nombre d'électrons de l'atome (chiffre à gauche);

2. Son poids atomique, le poids atomique de l'hydrogène étant pris conventionnellement égal à 1 (chiffre à droite).

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 1 H HYDROGÈNE																	2 4 He HÉLIUM
3 6,9 Li LITHIUM	4 9 Be BÉRYLLIUM											5 10,8 B BORE	6 12 C CARBONE	7 14 N AZOTE	8 16 O OXYGÈNE	9 19 F FLUOR	10 20,2 Ne NÉON
11 23 Na SODIUM	12 24,3 Mg MAGNÉSIUM											13 27 Al ALUMINIUM	14 28,1 Si SILICIUM	15 31 P PHOSPHORE	16 32,1 S SOUFRE	17 35,5 Cl CHLORE	18 39,9 A ARGON
19 39,1 K POTASSIUM	20 40,1 Ca CALCIUM	21 45 Sc SCANDIUM	22 47,9 Ti TITANE	23 51 V VANADIUM	24 52 Cr CHROME	25 54,9 Mn MANGANÈSE	26 55,8 Fe FER	27 58,9 Co COBALT	28 58,7 Ni NICKEL	29 63,5 Cu CUIVRE	30 65,4 Zn ZINC	31 69,7 Ga GALLIUM	32 72,6 Ge GERMANIUM	33 74,9 As ARSENIC	34 79 Se SÉLÉNIUM	35 79,9 Br BROME	36 83,8 Kr KRYPTON
37 85,5 Rb RUBIDIUM	38 87,6 Sr STRONTIUM	39 88,9 Y YTTRIUM	40 91,2 Zr ZIRCONIUM	41 92,9 Nb NIOBIUM	42 96 Mo MOLYBDÈNE	43 96 Tc TECHNÉTIUM	44 101,7 Ru RUTHÉNIUM	45 102,9 Rh RHODIUM	46 106,7 Pd PALLADIUM	47 107,9 Ag ARGENT	48 112,4 Cd CADMIUM	49 114,8 In INDIUM	50 118,7 Sn ÉTAIN	51 121,8 Sb ANTIMOINE	52 127,6 Te TELLURE	53 126,9 I IODE	54 131,3 Xe XÉNON
55 132,9 Cs CÆSIUM	56 137,4 Ba BARYUM	57 à 71 TERRES RARES SÉRIE DES LANTHANIDES	72 178,6 Hf HAFNIUM	73 180,9 Ta TANTALE	74 183,9 W TUNGSTÈNE	75 186,3 Re RHÉNIUM	76 190,2 Os OSMIUM	77 193,1 Ir IRIDIUM	78 195,2 Pt PLATINE	79 197,2 Au OR	80 200,6 Hg MERCURE	81 204,4 Tl THALLIUM	82 207,2 Pb PLOMB	83 209 Bi BISMUTH	84 210 Po POLONIUM	85 210 At ASTATE	86 222 Rn RADON
87 221 Fr FRANCIUM	88 226 Ra RADIUM	89 à 102 ÉLÉMENTS RARES SÉRIE DES ACTINIDES															

LANTHANIDES

57 138,9 La LANTHANE	58 140,1 Ce CÉRIUM	59 140,9 Pr PRASÉODYME	60 144,3 Nd NÉODYME	61 145 Pm PROMÉTHÉUM	62 150,4 Sm SAMARIUM	63 152 Eu EUROPIUM	64 156,9 Gd GADOLINIUM	65 159,2 Tb TERBIUM	66 162,5 Dy DYSPROSIUM	67 164,9 Ho HOLMIUM	68 167,2 Er ERBIUM	69 169,4 Tm THULIUM	70 173 Yb YTTERBIUM	71 175 Lu LUTÉCIUM

ACTINIDES

89 227 Ac ACTINIUM	90 232,1 Th THORIUM	91 231 Pa PROTACTINIUM	92 238,1 U URANIUM	93 237 Np NEPTUNIUM	94 242 Pu PLUTONIUM	95 243 Am AMÉRICIUM	96 243 Cm CURIUM	97 249 Bk BERKÉLIUM	98 249 Cf CALIFORNIUM	99 253 E EINSTEINIUM	100 255 Fm FERMIUM	101 256 Mv MENDÉLÉVIUM	102 No NOBÉLIUM

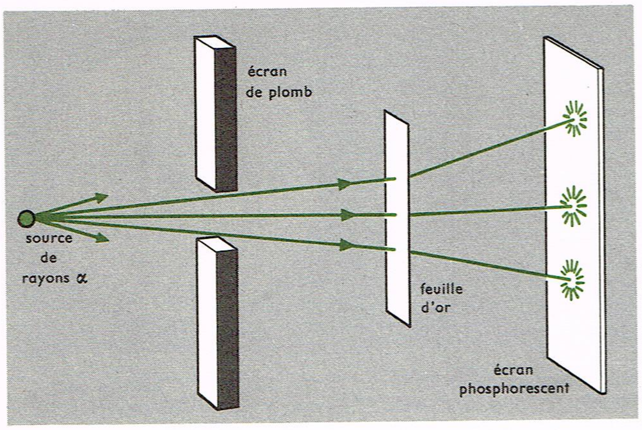
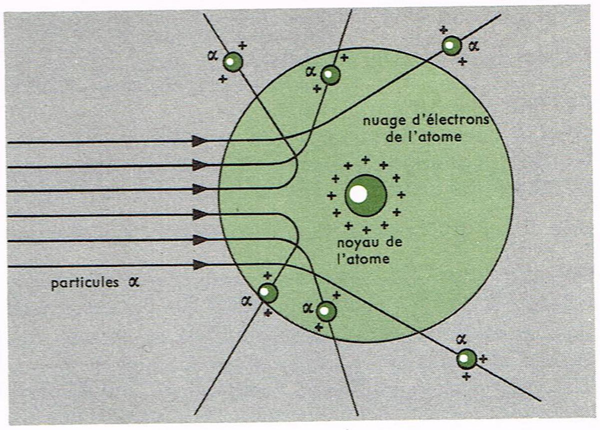


Schéma de l'expérience de Rutherford.

Schéma de la déviation des rayons α positifs par le noyau de charge positive.



La charge du noyau peut aussi être calculée, et des expériences sur les divers corps simples ont montré que le numéro atomique des chimistes correspondait exactement au nombre de charges positives du noyau, et donc au nombre d'électrons, puisque l'atome est électriquement neutre.

Le modèle de Rutherford nous fait déjà pénétrer dans la constitution de l'atome : un noyau central protégé et défendu par une couronne d'électrons.

Il est sans doute plus facile d'aborder la couronne, et l'on devine que les phénomènes électriques et chimiques devraient s'éclairer par la connaissance plus approfondie de cette couche superficielle.

En fait, la description complète de la nuée d'électrons entourant le noyau viendra essentiellement d'une étude détaillée de la lumière et de ses relations avec la matière. Ici s'ouvrira pour la physique une ère de découvertes paradoxales.

aspects corpusculaires du rayonnement électromagnétique : les quanta lumineux

Il suffit de regarder autour de soi pour se persuader qu'il doit y avoir une liaison étroite entre lumière et matière.

1° Qu'elle soit produite par un fil électrique porté à l'incandescence, par une bougie qui brûle ou par le Soleil lui-même, la lumière naît de la matière.

2° Qu'elle chauffe la surface de la Terre, fasse croître les plantes ou impressionne notre œil, la lumière est absorbée par la matière.

Comment se font ces mécanismes d'émission et d'absorption? Il est certain que la matière intervient ici dans toute sa complexité, que ces propriétés sont à mettre au compte de l'atome, de la molécule et des diverses liaisons moléculaires qui rendent différents les solides, les liquides et les gaz. Comment répondre autrement aux diverses questions telles que les suivantes :

Pourquoi la glace est-elle moins transparente que l'eau?

Pourquoi le cuivre est-il rouge, le carbonate de cuivre (CO_3Cu, ou vert-de-gris) vert, le sulfate de cuivre (SO_4Cu) bleu?

Pourquoi le cuivre, le carbonate de cuivre et le sulfate de cuivre donnent-ils à la flamme une même couleur verte?

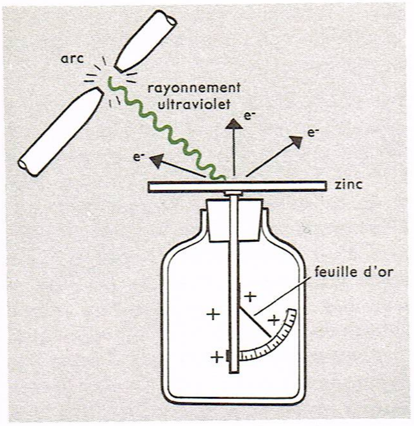
Ce dernier exemple montre en particulier que certains processus doivent être en relation directe avec les atomes; c'est à ces processus atomiques que nous nous intéresserons ici.

Que sait-on jusqu'à présent de la lumière? La lumière est une onde électromagnétique qui se propage dans le vide à la vitesse de 300 000 km/s; cette onde est caractérisée par sa longueur d'onde λ et par son intensité I (qui mesure l'énergie transportée par l'onde). La lumière « visible » n'est finalement qu'une toute petite partie du spectre des ondes électromagnétiques différant entre elles seulement par leur longueur d'onde, spectre qui va des ondes de radio aux rayons γ, en passant par les rayons infrarouges, la lumière « visible », les rayons ultraviolets et les rayons X. Il y a ainsi une identité foncière entre toutes ces ondes et pourtant leurs effets sur la matière sont essentiellement différents. On peut le comprendre au moins qualitativement de la façon suivante : **l'onde électromagnétique est absorbée par la matière par l'intermédiaire de petits résonateurs qui peuvent osciller à la même fréquence qu'elle.** Pour expliquer les propriétés des différentes ondes, on suppose que ces résonateurs sont de plusieurs sortes, de fréquences déterminées. Ainsi notre œil est un résonateur adapté seulement à la lumière « visible »; l'infrarouge favorise l'agitation des molécules et se transforme surtout en chaleur; l'ultraviolet favorise les réactions chimiques telles que le brunissement de la peau en haute montagne; les rayons X traversent la matière (sauf les matières lourdes comme le plomb) sans trouver de résonateur à leur convenance.

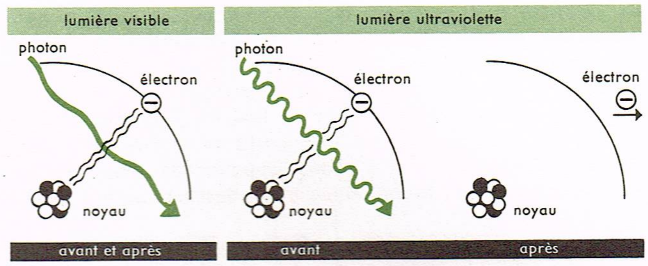
Cependant, à l'échelle de l'atome, cette conception (classique) de la lumière se révélera insuffisante et devra faire place à une autre, plus corpusculaire et discontinue, à l'image de l'atome lui-même. Tout se passe alors comme si la lumière mettait « en veilleuse » sa nature ondulatoire pour se montrer sous un aspect corpusculaire. Laissons de côté la question de savoir si ces deux aspects peuvent coexister pacifiquement; nous nous bornerons ici à constater qu'il est impossible d'éviter ce dualisme. Le cas sera particulièrement net pour les phénomènes d'absorption de la lumière à l'échelle atomique; ainsi l'effet photo-électrique, incompréhensible pour une onde, devient clair et simple pour un corpuscule.

l'effet photo-électrique

La lumière est capable d'arracher des électrons aux atomes et, par le fait même, de donner naissance à un courant électrique; c'est en cela que consiste l'effet photo-électrique, dont la « cellule » — chère aux photographes — est loin d'être la seule application pratique.



Première mise en évidence de l'effet photo-électrique : l'arc électrique émet des rayons ultraviolets qui sont dirigés sur une plaque de zinc reliée à un électroscope. La feuille d'or, très légère, en liaison avec la tige, se sépare d'elle par répulsion électrostatique indiquant que l'appareil s'est chargé électriquement.



Le photon de lumière visible, trop mou, peut heurter un atome, mais il le laisse intact; le photon de lumière ultraviolette, plus énergique, arrive à casser la liaison électron-atome et peut disparaître en cédant toute son énergie.

Un appareillage assez simple permit dès 1890 de mettre en évidence cet effet : sous l'action de rayons ultraviolets, le zinc se charge électriquement (plus exactement, il se charge en perdant des électrons, comme on arriva à le montrer après la découverte de ces mêmes électrons).

De multiples expériences, précises et complètes, conduisirent aux conclusions suivantes, valables pour un corps donné (le zinc, par exemple) :

1° Pour une longueur d'onde fixe de la lumière, la vitesse d'éjection des électrons ne dépasse pas une certaine valeur, quelle que soit l'intensité de la lumière;

2° Il y a une longueur d'onde de la lumière au-dessus de laquelle l'effet photo-électrique ne se manifeste plus; il y a donc un « seuil » de l'effet photo-électrique. Si le zinc, par exemple, émet des électrons avec la lumière ultraviolette, il ne donne rien avec la lumière visible.

Essayons maintenant d'interpréter ces résultats en imaginant comment peut se faire l'action de la lumière sur les électrons.

Dans le cadre de la théorie classique des ondes électromagnétiques, la première propriété est absolument incompréhensible. En effet, si la lumière est semblable à une onde, en augmentant l'intensité de cette onde lumineuse, nous devrions augmenter la vitesse des électrons, tout comme à la mer, où nous ressentons bien plus l'effet des grosses vagues que celui des petites! La question peut être abordée sous différents aspects, la contradiction reste toujours aussi flagrante.

Quant à la deuxième propriété, l'existence d'un seuil, nous pouvons peut-être supposer que l'électron, pour sortir de l'atome, doit osciller à une certaine fréquence, bien que nous n'en saisissions pas la raison profonde.

Mais laissons là ces contradictions et ces obscurités que présente la théorie classique et faisons table rase d'idées *a priori*.

Des deux propriétés indiquées ci-dessus, nous pouvons nous faire alors une idée simple, et qui finalement n'est pas simpliste, mais juste, en nous disant : **la lumière est formée de petits grains d'énergie fixe et ce sont ces grains (nous les appellerons des « photons ») qui interagissent directement avec les électrons.** Un électron est éjecté par un seul photon et ne peut prendre que l'énergie cinétique que celui-ci est susceptible de lui donner. Si j'augmente l'intensité lumineuse, j'augmente le nombre de photons, mais non leur énergie propre, et, partant, j'augmente le nombre des électrons libérés, mais non leur énergie cinétique. La vitesse maximale des électrons se trouve ainsi facilement expliquée.

Pour comprendre l'existence assez mystérieuse du seuil photo-électrique, il suffit de supposer maintenant que l'énergie propre des photons est d'autant plus grande que leur longueur d'onde est petite (en ce sens, les rayons X ou ultraviolets seront plus énergiques que les rayons lumineux visibles). En effet, pour libérer l'électron de l'atome, il faut certainement apporter une certaine énergie et seuls en sont alors capables certains photons aux longueurs d'onde petites.

Evidemment, $E = \frac{\text{constante}}{\lambda}$ est l'expression la plus simple possible pour une énergie E des photons de longueur d'onde λ satisfaisant à cette condition. Pour des raisons de commodité, nous pouvons écrire $[E = h\nu]$, où h est une constante et $\nu = \frac{c}{\lambda}$ la fréquence du photon.

Voilà le raisonnement simple qu'Einstein a osé faire le premier en 1905 (ce qui lui a valu le prix Nobel), guidé d'ailleurs en cela par les travaux de Planck, qui introduisit le premier cette constante h qui porte son nom.

comment la cellule photo-électrique a permis de confirmer la théorie d'Einstein

Voyons de plus près comment ces idées ont pu être vérifiées, et les propriétés de l'effet photo-électrique bien établies. La cellule photo-électrique dite « photo-émissive » (nous verrons plus loin qu'il y a d'autres types de cellules, dont dérive par exemple la cellule des photographes) a été d'un grand secours pour cela. Elle est essentiellement constituée par une ampoule de verre contenant une électrode A sensible à la lumière et une autre électrode B; le vide règne à l'intérieur de l'ampoule, et les électrodes, reliées à un circuit électrique extérieur, sont soumises à une différence de potentiel. La cellule ne fournit

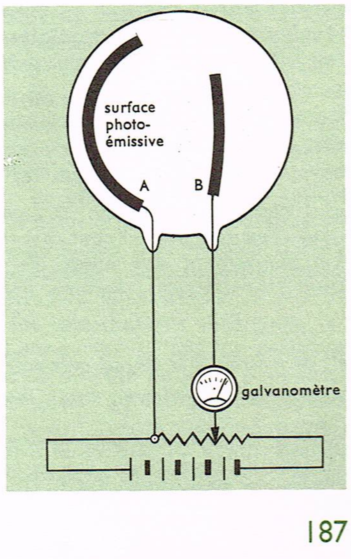


Schéma de montage d'une cellule photo-émissive.

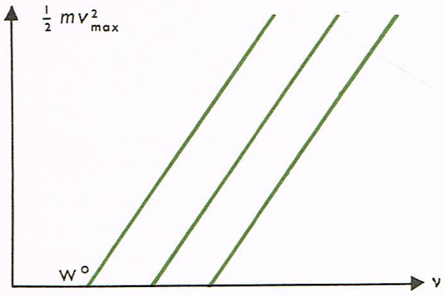
de courant que si la lumière tombe sur la surface sensible. Selon les idées d'Einstein, que se passe-t-il? Un photon de la lumière monochromatique d'énergie $h\nu$, en frappant la plaque, a disparu complètement en donnant toute son énergie à un électron; cette énergie a servi à libérer l'électron de l'atome en lui donnant également une certaine vitesse. Les électrons les moins liés à l'atome (nous appellerons W_0 leur énergie de liaison) seront chassés avec une vitesse v_{max} et une énergie cinétique correspondante $\frac{1}{2} mv^2_{max}$ telle que nous ayons

$$h\nu = W_0 + \frac{1}{2} mv^2_{max}.$$

Que cette vitesse maximale existe réellement, on peut le vérifier d'une façon simple. Si l'électrode B est à un potentiel négatif par rapport à celui de la surface photosensible A, les électrons sont repoussés par cette électrode B; mais si la différence de potentiel est faible, leur vitesse initiale leur permet d'atteindre quand même cette dernière. Si l'on augmente cette différence de potentiel, il y a un moment où les électrons ne sont plus assez rapides pour arriver jusqu'à l'électrode B : le courant doit donc cesser. C'est ce qui arrive en effet.

Augmenter l'intensité lumineuse ne sert à rien, ce qui montre bien que la vitesse maximale n'en dépend pas.

On a d'ailleurs ainsi un moyen de mesurer cette vitesse maximale (1), et par là même de vérifier la formule d'Einstein $h\nu = W_0 + \frac{1}{2} mv^2_{max}$, en faisant varier la longueur d'onde de



L'énergie cinétique maximale des photo-électrons varie linéairement en fonction de la fréquence de la lumière incidente. Les droites ont toujours même pente (h), mais sont décalées suivant le métal utilisé (seuil photo-électrique W_0 différent).

la lumière monochromatique incidente. L'expérience a, en effet, montré que les courbes donnant l'énergie cinétique $\frac{1}{2} mv^2_{max}$ en fonction de ν sont des droites, et ces droites permettent d'ailleurs de mesurer aussi bien la constante h que l'énergie de liaison W_0 (valeur de $h\nu$ pour laquelle $v_m = 0$ au seuil de l'effet photo-électrique).

Il se trouve que cette théorie du photon lumineux est amplement vérifiée par le fait que les constantes h et W_0 ont été trouvées identiques dans l'interprétation d'autres phénomènes tout à fait différents de l'effet photo-électrique.

1° La constante h mesurée est celle-là même qui a été introduite en 1899 par le physicien allemand Planck pour expliquer le comportement bizarre des corps dits « noirs », c'est-à-dire des corps qui absorbent toutes les radiations qu'ils reçoivent.

2° L'énergie de liaison W_0 est la même que celle que l'on peut mesurer dans l'effet thermo-ionique qui est à la base du fonctionnement des tubes électroniques d'un poste de radio (non transistorisé); comme une eau qui bout fait jaillir à sa surface des gouttelettes qui retombent ensuite, un filament incandescent libère des électrons qui peuvent être ensuite recueillis par une électrode positive.

Ainsi, pour expliquer l'effet photo-électrique, il faut se faire de la lumière la représentation suivante : la lumière est formée de grains, ou photons, d'énergie propre $E = h\nu$. La constante h est très petite ($h = 4{,}1 . 10^{-15}$ électron-volt/seconde) : une surface de 1 cm² éclairée par la lumière d'une bougie placée à 1 mètre reçoit environ 10^{14} photons par seconde. Néanmoins, nous sommes ici en présence d'un phénomène discontinu : on ne peut pas avoir de quantité d'énergie lumineuse plus petite que la quantité $h\nu$ (de même que l'on ne peut avoir ni masse ni charge inférieures à celles de l'électron). C'est pourquoi on appelle cette quantité $h\nu$ un « quantum » lumineux. La théorie classique des ondes électromagnétiques ne prévoit pas l'existence de tels quanta et il faudra inventer une nouvelle théorie.

Elle le sera dans les années 1920-1930 sous l'impulsion de Louis de Broglie et d'Heisenberg.

Si l'effet photo-électrique a eu des applications théoriques importantes, comme on vient de le voir, son importance pratique n'a cessé de croître et il est intéressant ici d'en voir quelques applications.

cellules photo-émissives

Il est évident que les cellules photo-émissives, dont le principe a déjà été décrit plus haut, ne pourront être d'utilité courante que si elles sont sensibles à la lumière visible. Malheureusement, le seuil photo-électrique de la plupart des corps, des métaux en particulier, se situe dans l'ultraviolet. Quelques métaux seulement sont sensibles à la lumière visible : ce sont ceux chez lesquels l'énergie de liaison des électrons aux atomes est faible, c'est-à-dire, par le fait même, ceux dont la réactivité chimique est grande. Les métaux tels que le sodium, le potassium, le césium, qui s'oxydent spontanément à l'air, sont de ceux-là. Aussi la préparation de la surface photo-sensible — souvent en césium — demande-t-elle des précautions spéciales, toutes les opérations devant être faites sous vide.

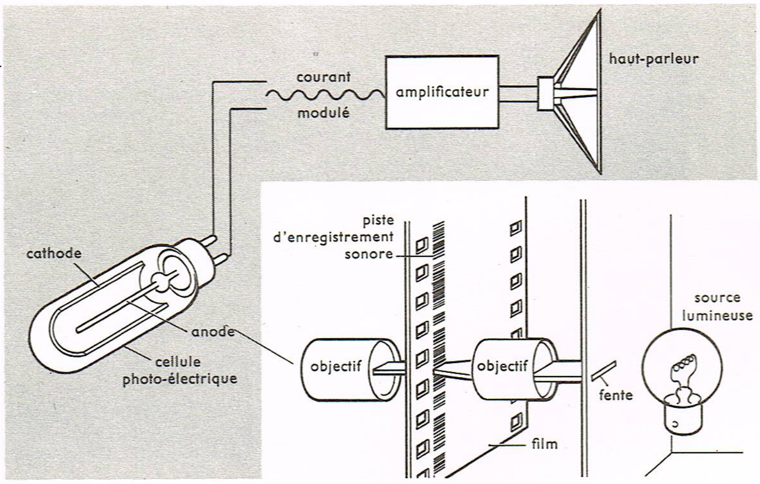
Malgré les techniques poussées et les perfectionnements apportés aux surfaces photo-sensibles, le courant émis par une cellule photo-émissive reste faible. L'inconvénient n'est pourtant pas majeur, car on peut amplifier ce courant et détecter ainsi de très faibles intensités de lumière : à 1 km de distance, par exemple, la lueur d'une bougie peut être facilement décelée. Par ailleurs, la cellule photo-électrique est capable de transcrire fidèlement et instantanément les variations d'intensité lumineuse par des variations du courant photo-électrique correspondant.

On perçoit tout le parti qu'on peut tirer de si remarquables caractéristiques.

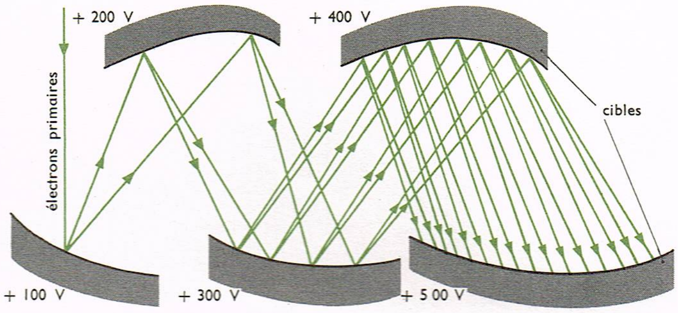
Dans toutes les applications, le courant photo-électrique doit être amplifié, et il peut l'être par des amplificateurs de type courant; mais au lieu d'employer un amplificateur séparé, on réalise maintenant des cellules photo-électriques qui sont en même temps amplificatrices. On leur donne le nom de « **photomultiplicateurs** ». Le principe de fonctionnement est le suivant. La lumière tombe sur la cathode et en extrait des électrons par effet photo-électrique. Accélérés par une première anode portée par exemple à 100 V, ces électrons viennent la frapper et entraînent, grâce à la vitesse acquise, un nombre d'électrons beaucoup plus grand. Ces électrons secondaires se dirigent vers une deuxième anode, portée par exemple à 200 V, et le processus peut se continuer jusqu'à une anode finale. Il y a eu ainsi multiplication du nombre d'électrons initialement émis par la lumière (d'où ce nom de « photomultiplicateur »), multiplication qui peut aller jusqu'à un rapport de 100 000 pour 1 et qui est d'ailleurs réglable si l'on fait varier le potentiel des anodes.

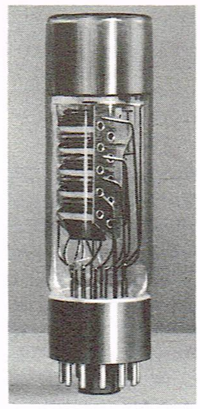
---

(1) Au moment où le courant cesse pour une valeur V_0 du potentiel, toute l'énergie cinétique des électrons $\frac{1}{2} mv^2_{max}$ se transforme en énergie électrostatique eV_0, et l'on a $eV_0 = \frac{1}{2} mv^2_{max}$. Une unité commode d'énergie est ici l'électron-volt (eV), qui correspond à l'énergie cinétique que prend un électron accéléré par une différence de potentiel de 1 volt.



Reproduction du son en cinématographie.
A l'enregistrement du son, les vibrations sont transformées en variations de courant électrique, qui, elles-mêmes, font varier l'intensité d'une source lumineuse par un procédé qui suit fidèlement les variations du courant; le film « sonore » est impressionné par cette lumière et son opacité variable se trouve ainsi modulée sur le son. A la projection, une forte lumière éclaire le film et l'intensité lumineuse transmise derrière le film est « lue » par une cellule photo-électrique; celle-ci donne naissance à un courant modulé qui, après amplification, agit sur le haut-parleur.





Tube photomultiplicateur et schéma de principe.



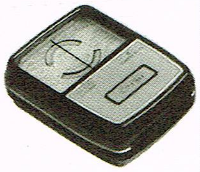
La télévision est un autre domaine qui demande des appareillages à la fois rapides et précis. Les cellules photo-électriques en constituent un des éléments de base. A l'émission, elles transforment les variations de lumière de la scène en modulation des ondes électromagnétiques qui se répandent dans toute une région et, même maintenant avec les satellites artificiels, de par le monde entier. A la réception, le courant photo-électrique plus ou moins intense donne un flux d'électrons qui impressionne plus ou moins l'écran devant lequel vous êtes assis. Le problème est évidemment compliqué du fait qu'il faut retransmettre non pas seulement un point lumineux donné, mais toute une image; il faut donc un système de balayage ou de lecture rapide de la scène à l'émission et de l'écran à la réception.

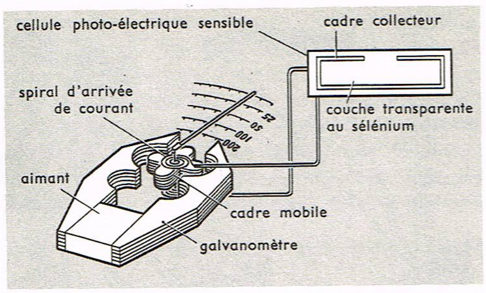
Le photo-multiplicateur a l'avantage sur les amplificateurs à lampes ordinaires de donner une amplification quasi instantanée; sa limite de résolution dans le temps correspond en effet au temps de transit des électrons, rendu petit par l'encombrement minimal des électrodes (il est de l'ordre de 10^{-8} à 10^{-9} seconde). Il s'adapte ainsi très bien aux courants de haute fréquence et reçoit une utilisation courante dans la physique actuelle pour la détection et le comptage soit des photons, soit de toutes les autres particules; dans ce dernier cas, les particules viennent d'abord frapper un « scintillateur », matière fluorescente capable d'émettre de la lumière, que détecte le photo-multiplicateur.

cellules photo-électriques et photopiles

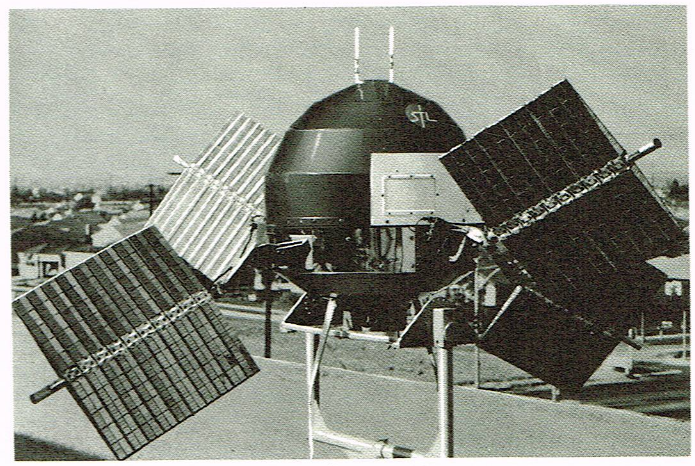
Les applications citées ci-dessus demandent de la précision et de la fidélité, mais on peut en imaginer d'autres qui ne requièrent pas tant de qualités. Si vous possédiez un hôtel au bord d'une route nationale et que vous vouliez attirer l'attention de l'automobiliste nocturne, vous pourriez utiliser évidemment une enseigne lumineuse; l'effet psychologique serait encore meilleur si, à son passage, vous allumiez cette enseigne : vous pourriez le faire en utilisant une cellule photo-électrique commandée par les phares de l'automobile; mais il suffirait d'une cellule qui agît en « tout ou rien » sur la commande électrique de l'enseigne. Point n'eût été besoin de mesurer l'intensité des phares! Même pour la photographie, une cellule photo-électrique à hautes performances n'est pas nécessaire puisqu'on réussit à prendre d'excellentes photos sans cellule. On utilise toujours pour cela des cellules à **semi-conducteurs,** qui sont robustes, simples, utilisables sans source auxiliaire de courant, et qui fournissent des informations suffisantes sur l'intensité lumineuse. Les semi-conducteurs, tels le sélénium et le germanium, sont ainsi appelés à cause de leurs propriétés électriques étranges : ils laissent passer le courant plus facilement dans un sens que dans l'autre; propriété mal connue encore, mais qui doit se localiser à la surface de séparation du semi-conducteur et du métal ordinaire. Sous l'action de la

Doc. Thomson-Houston.





Posemètre à photopile.



Certains satellites ont près de 10 000 photopiles disposées de façon à recevoir dans les meilleures conditions la lumière solaire.

Satellite « Explorer VI ».

lumière, ils sont capables, par un effet photo-électrique interne, de donner naissance à un courant dans un sens déterminé. Une cellule de photographe est ainsi un « générateur » d'électricité au rendement assez faible, pourtant suffisant pour faire dévier l'aiguille d'un microampèremètre. Nous avons déjà affaire à une « photopile », capable de transformer la lumière en électricité. On a étudié toutes sortes de semi-conducteurs pour améliorer le rendement électrique, qui est de toute façon toujours supérieur à celui d'une cellule photo-émissive; on arrive maintenant à réaliser des photopiles capables de débiter un courant appréciable. Les batteries d'accus des satellites interplanétaires sont rechargées par des photopiles qui utilisent directement l'énergie solaire.

Ces diverses applications donnent un petit aperçu de l'importance pratique qu'a pris l'effet photo-électrique.

effet Compton

L'effet photo-électrique, en dehors de son importance pratique, a permis de comprendre certains mécanismes de transformation d'énergie à l'échelle atomique : il faut pour cela se représenter la lumière comme formée de photons ou quanta lumineux, sortes de petites boules d'énergie qui sont indépendantes les unes des autres et qui se suivent à une cadence de mitrailleuse ultra-rapide.

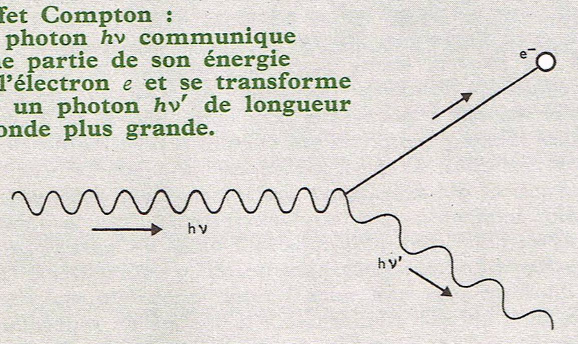
Si dans l'effet photo-électrique ces petites boules disparaissent en cédant toute leur énergie aux électrons, ne peut-on pas imaginer qu'elles puissent ricocher sur des atomes ou des électrons (c'est-à-dire sur d'autres boules) en perdant seulement une partie de leur énergie? Le phénomène serait alors identique à celui de la collision de deux boules de billard, donnant lieu à un transfert d'énergie. La longueur d'onde des photons, directement liée à leur énergie ($E = h\nu$), devrait donc être plus grande après diffusion qu'avant. La théorie classique de la diffusion des ondes électromagnétiques ne prévoit pas du tout un tel changement. On comprend ainsi l'effort que fit Compton vers les années 1920 pour mettre en évidence cet effet, à un moment où l'on se posait encore beaucoup de questions sur la nature du rayonnement lumineux.

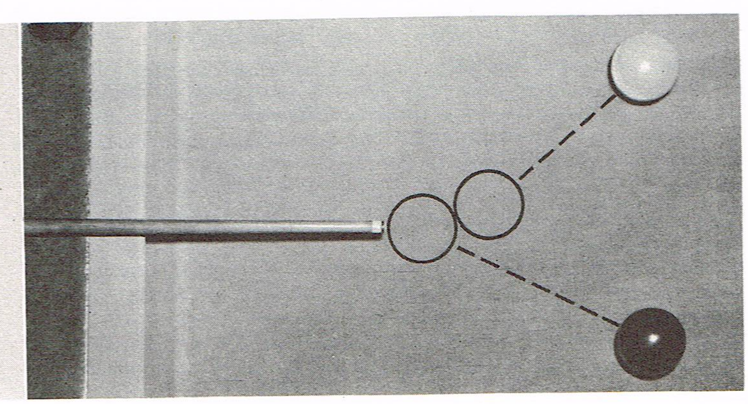
Si tout se passe comme au jeu de boules, on peut se rendre compte que la collision avec un atome donnera un effet très faible et indécelable : l'atome, qui est lourd (relativement bien sûr), prendra très peu de recul à cause de son inertie, et le changement d'énergie (ou de longueur d'onde) du photon sera très petit. Au contraire, la collision avec un électron, au moins 2 000 fois plus léger, pourra donner un effet appréciable, à condition toutefois que le photon ait une énergie suffisante pour l'arracher de l'atome. Nous savons par l'effet photo-électrique qu'il faut pour cela des quanta très énergiques; aussi Compton fit-il ses expériences avec des rayons X durs. (On savait à cette époque produire des rayons X, découverts en 1895 par Röntgen, et bien mesurer leurs longueurs d'onde.)

Avec de tels rayons X, l'énergie de liaison de l'électron est faible devant l'énergie du quantum $h\nu$, si bien que tout doit se passer comme si la collision se faisait avec un électron libre, et le phénomène devient très simple à prévoir : les rayons X les plus déviés seront ceux qui auront perdu le plus d'énergie et, suivant leur déviation, les nouvelles longueurs d'onde peuvent être exactement prédites. Les mesures de Compton ont montré, en effet, qu'il existait bien deux types de diffusions : une sans changement de longueur d'onde, prévue par la théorie classique, et l'autre avec changement de longueur d'onde, répondant exactement à l'exemple du jeu de boules.

Ainsi la structure discontinue de la lumière et des ondes électromagnétiques s'est imposée avec de plus en plus de certitude, à l'image de la structure discontinue de la matière elle-même. On découvrit d'ailleurs plus tard que les rayons γ, provenant

Effet Compton : le photon $h\nu$ communique une partie de son énergie à l'électron e et se transforme en un photon $h\nu'$ de longueur d'onde plus grande.





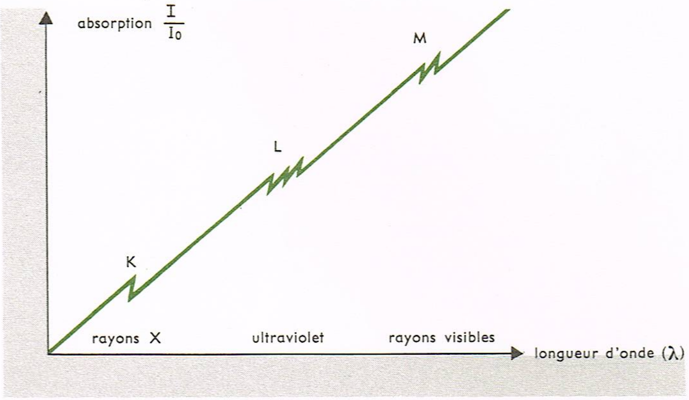
de la radio-activité du noyau atomique, plus énergiques encore que les rayons X, pouvaient donner naissance à une paire d'électrons (voir la seconde partie de ce chapitre), c'est-à-dire se transformer en matière; ce qui vint confirmer encore la ressemblance entre les photons et les autres particules de la matière.

l'énergie qui lie les électrons aux noyaux

Après avoir bien vu comment se comportait la lumière à l'échelle atomique, nous sommes à même de mieux comprendre comment elle peut nous renseigner sur la structure de l'atome. L'effet photo-électrique lui-même donne ici sa contribution, par l'intermédiaire de son seuil, qui, nous l'avons vu, nous permet de connaître l'énergie de liaison des électrons les moins liés des divers atomes.

Ce seuil est en général dans le rayonnement ultraviolet, ce qui donne pour l'électron une énergie de liaison de l'ordre de 2 à 13 eV. Si l'on étudie les rayons X — plus énergiques que les rayons ultraviolets — et leur absorption par la matière, on obtient d'autres renseignements qui donnent une première indication sur l'arrangement des électrons dans les atomes.

En effet, les rayons X traversent assez facilement la matière tout en étant partiellement absorbés : les recevant sur une plaque photographique, on peut, connaissant l'intensité émise et mesurant l'intensité reçue, en déduire l'intensité absorbée. En changeant la longueur d'onde des rayons X, on peut connaître la variation de l'absorption en fonction de la longueur d'onde. Les courbes obtenues ont la même forme pour tous les corps simples : plus λ est grand, plus il y a absorption; cela est bien conforme au fait que beaucoup de corps sont opaques à la lumière visible, mais transparents aux rayons X. De plus — et ceci est important —, il y a des discontinuités caractéristiques dans la courbe : ainsi l'aluminium présente des discontinuités dans l'ultraviolet, coïncidant exactement avec ce que nous avons appelé plus haut « seuil de l'effet photo-électrique ». Quand on diminue la longueur d'onde, l'absorption diminue aussi, jusqu'au moment où se produit une augmentation brusque, qu'on interprète comme un nouveau seuil de l'effet photo-électrique pour des électrons plus liés que les précédents. On observe ainsi que les électrons de l'aluminium semblent se diviser en trois groupes, K, L, M, assez différents par leur énergie de liaison. On dira qu'il y a plusieurs couches d'électrons autour du noyau central, pour des raisons qui deviendront très claires plus loin. Comment se comportent exactement les électrons autour du noyau atomique est évidemment une question intéressante à se poser. Les renseignements nous seront donnés encore une fois par l'étude de la lumière, et plus précisément par le phénomène de son émission par la matière.



Courbe d'absorption des rayons X et ultraviolets dans l'aluminium.

de la spectroscopie à l'édifice électronique des atomes

La première analyse spectroscopique remonte probablement à Newton (XVII[e] siècle), quand il montra que la lumière solaire était un mélange de lumières de couleur, reproduisant exactement le spectre de l'arc-en-ciel. Il le fit en utilisant simplement un prisme, à travers lequel la lumière se réfracte différemment suivant les longueurs d'onde. On détient là le principe d'un type de spectroscope qui servit à établir des centaines de spectres et à en mesurer les diverses longueurs d'onde. Au lieu de laisser tomber sur le prisme toute l'intensité lumineuse, on la limite par une fente, qui, grâce à une première lentille, envoie sur le prisme un faisceau de lumière parallèle et, grâce à une deuxième, donne une image qu'on peut recevoir sur une plaque photographique ou regarder avec une simple loupe, ou, pour être plus juste, avec un oculaire de lunette. Si la lumière est « monochromatique », on obtient une image bien définie de la fente, qui apparaît ainsi comme une raie correspondant à la longueur d'onde précise de la lumière émise. Si elle est complexe, on obtient soit une suite de raies, soit un spectre continu.

Malgré cette simplicité d'appareillage, l'histoire de la spectroscopie n'a commencé vraiment qu'au XIX[e] siècle, avec Fraunhofer, Kirchhoff, Bunsen. On observa d'abord que le spectre du Soleil était formé d'une suite de raies discontinues, les unes plus brillantes que les autres. On se mit ensuite à observer le spectre des flammes, y plongeant différents corps chimiques, qui donnaient de belles colorations rouges, vertes, orangées, etc. On s'aperçut ainsi que ces corps donnaient des raies spectrales caractéristiques. On se mit à les classer, découvrant même de nouveaux éléments chimiques inconnus jusqu'alors. On se mit aussi à étudier les spectres donnés par des arcs électriques ou par des étincelles qu'on pouvait faire jaillir entre les électrodes d'une bobine de Ruhmkorff. En changeant la nature des électrodes, on étudiait les spectres de divers métaux. Bientôt, on eut un nouveau moyen d'investigation avec les décharges électriques dans les gaz raréfiés : un tube allongé contenant un gaz dont la pression est réduite à quelques millimètres de mercure s'illumine sous une tension électrique de quelques milliers de volts entre les électrodes, en rouge si le gaz est du néon, en bleu si c'est du crypton, etc.

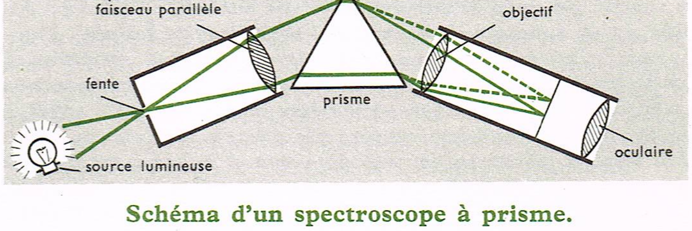
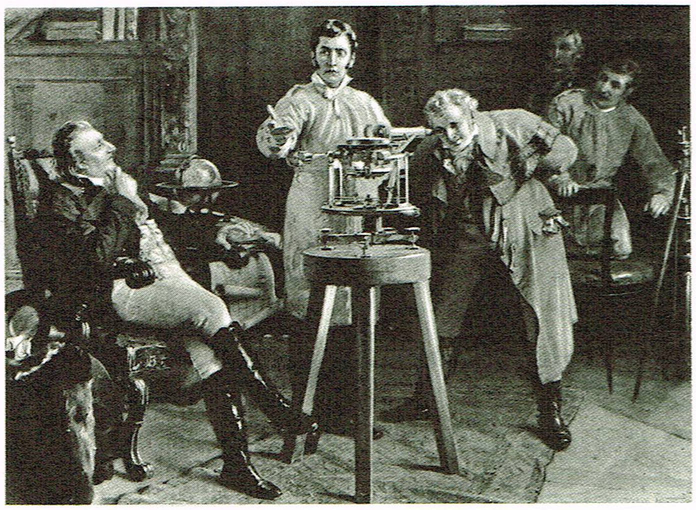


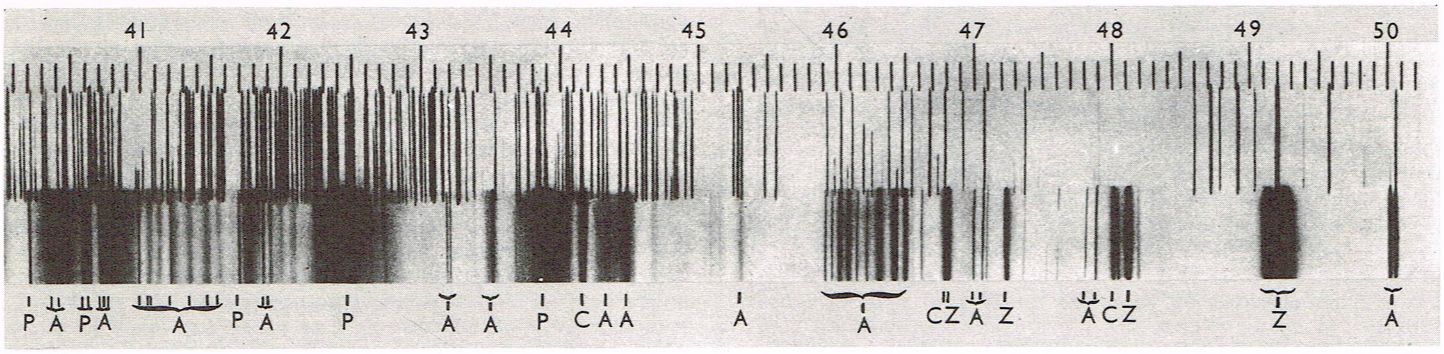
Schéma d'un spectroscope à prisme.



Joseph Fraunhofer présente son spectromètre (d'après un tableau de R. Wimmer).

A l'aube du XIX[e] siècle, au moment où Auguste Comte affirmait que jamais l'homme ne pourrait savoir ce qui se passe à l'intérieur d'une étoile, Fraunhofer étudie le spectre solaire, découvre les raies du fer, en déduit la présence de fer dans l'atmosphère solaire et invente la technique (spectroscopie) qui allait permettre rapidement de connaître le centre du Soleil mieux que le sous-sol terrestre.

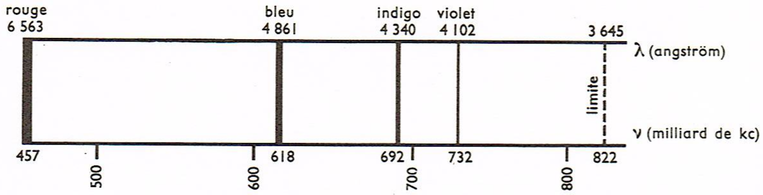
Ci-dessous, **une partie du spectre du fer (partie supérieure) comparée au spectre d'un alliage plomb (P), zinc (Z) et cadmium (C). Les raies A sont celles de l'air.**



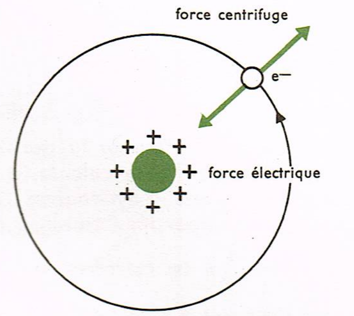
Le spectre de la lumière visible fut évidemment le premier à être étudié. Pour les parties invisibles, il fallut attendre des techniques plus élaborées de spectrographie. Les rayons infrarouges sont maintenant détectés par des récepteurs thermiques, des cellules photo-électriques ou des plaques photographiques spéciales; les rayons ultraviolets le sont également par des plaques photographiques, mais, en outre, le verre étant alors très absorbant, il faut souvent utiliser des prismes en quartz. Quant aux rayons X, découverts en 1895, ils ne sont guère déviés par le prisme, qui, alors, n'est plus d'aucune utilité. Il fallut attendre 1913 et von Laue pour savoir mesurer leur longueur d'onde grâce à leur diffraction par le réseau très fin que forme un cristal; de tels spectroscopes à réseaux étaient déjà connus dans la spectroscopie de la lumière ordinaire.

Qu'elle soit due à l'agitation thermique produite directement par la chaleur d'une flamme ou celle d'un corps porté au rouge, qu'elle soit le résultat d'une décharge électrique dans les gaz raréfiés ou dans l'air, il faut conclure que l'émission de la lumière est due à une certaine « excitation » des molécules ou des atomes de la matière. On devine pourtant les difficultés qu'il a fallu surmonter pour établir, dans un spectre complexe et confus, ce qui était à mettre au compte de la molécule, de l'atome, neutre ou ionisé, sinon de l'élément lui-même. Un grand pas fut fait quand on acquit la conviction que, des deux types de spectres rencontrés, le « spectre de bandes », continu, et le « spectre de raies », discontinu, le premier était dû à l'excitation de la molécule, le second à l'excitation de l'atome. Finalement, c'est plus de cinquante ans d'un travail de patience et de précision qu'il a fallu pour établir les principaux spectres avec leurs caractéristiques détaillées. Mais c'est aussi grâce à ce travail laborieux que, en plus des progrès énormes qu'elle a fait faire à la compréhension de l'atome, la spectroscopie est devenue d'un emploi courant dans le monde d'aujourd'hui. La chimie lui doit l'un de ses plus puissants moyens d'analyse. Des proportions infimes d'un élément peuvent être détectées dans un mélange, même s'il s'agit de l'arsenic, qui ne fait pourtant pas souvent l'unanimité des experts. Une analyse quantitative par mesure de l'intensité lumineuse du spectre donne aussi de très bons résultats et se pratique ainsi dans le contrôle de fabrication des alliages. La chimie organique s'adresse surtout aux spectres de bandes qui fournissent d'irremplaçables renseignements sur les molécules, en particulier au spectre par absorption par lequel on analyse la lumière qui a traversé d'abord l'échantillon étudié. On a établi les spectres d'absorption des différentes liaisons chimiques, ce qui est d'un grand secours pour définir des formules chimiques aussi compliquées que celles des hormones. L'industrie des pétroles emploie systématiquement la spectrographie infrarouge par absorption, pour déterminer les proportions des différents hydrocarbures dans un nouveau gisement, et pour contrôler les opérations de cracking et de raffinage.

L'astronomie a constamment recours à la spectroscopie. Si l'on sait que le Soleil contient beaucoup d'hydrogène et d'hélium, c'est bien à elle qu'on le doit. L'hélium a même été découvert dans le Soleil avant de l'être sur la Terre. On a même cru un moment avoir découvert dans le Soleil un nouvel élément inconnu sur Terre; on a pu finalement montrer que le spectre récalcitrant était celui du fer treize fois ionisé (atome de fer ayant perdu treize électrons), spectre inconnu au laboratoire, dont les moyens ne permettent pas encore de reproduire les conditions qui règnent dans le Soleil. Les températures superficielles des étoiles se déduisent de leur couleur et de leur spectre, notre Soleil se plaçant dans une moyenne honnête avec ses 6 000 °C. On a pu, toujours grâce à la spectroscopie, observer un déplacement vers le rouge des raies extra-galactiques et, à cause de l'effet Doppler-Fizeau, parler de la vertigineuse fuite de galaxies s'éloignant à des vitesses comparables à celle de la lumière. Sur cette grande vision a été bâtie une théorie de l'expansion de l'univers (G. Lemaître, 1927).



Série des raies de Balmer rangées dans une échelle de fréquences (γ) équidistantes $\left(\gamma = \frac{c}{\lambda}\,;\ 1\,\text{Å} = 10^{-4}\,\mu\right)$.



Modèle d'atome : si l'électron tourne autour du noyau, la force centrifuge qui apparaît peut empêcher l'électron de tomber sur le noyau.

Toutes ces applications n'auraient pas été possibles, on le devine, avec une simple collection de spectres réunis en un recueil de recettes de cuisine. **Il a fallu auparavant interpréter, classer, comprendre enfin.** Difficile était le travail, car, au départ, il s'agissait d'un véritable puzzle, dont devaient finalement sortir des découvertes sensationnelles et révolutionnaires sur le monde de l'atome.

le spectre de l'hydrogène et l'atome de Bohr

Le premier homme à mettre de l'ordre dans la maison et à systématiser les connaissances fut le Suisse Balmer. Il s'attaqua à l'hydrogène — le plus simple de tous les éléments — en produisant son spectre grâce à une décharge électrique dans un tube contenant ce gaz. A côté d'un spectre continu dû à la molécule, il trouva quatre raies fines dans la lumière visible (une rouge, une bleue, une indigo, une violette) et plusieurs autres, de plus en plus rapprochées, dans l'ultraviolet; il les attribua à l'atome d'hydrogène. Après plus d'un essai infructueux, il réussit à relier ces raies par une formule simple :

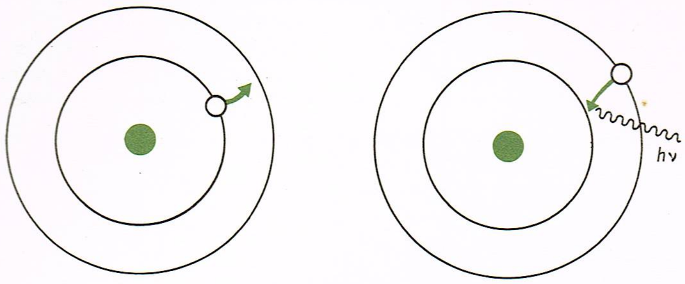
$$\nu = R\left(\frac{1}{2^2} - \frac{1}{n^2}\right),$$

dans laquelle il donnait à n des valeurs entières 3, 4, 5, ..., et ajustait la constante R pour retrouver la première raie.

De 1885, date à laquelle Balmer donna sa formule, à 1913, date à laquelle Bohr en trouva une explication, il fallut attendre un quart de siècle, mais l'explication fit l'effet d'une bombe dans les milieux scientifiques. D'ailleurs, pendant ce temps, les physiciens avaient été loin d'être inactifs, traçant la route à suivre. Planck, en 1899, avait fait une première révolution en faisant accréditer la notion de discontinu et de quantum dans les échanges d'énergie; Einstein, en 1905, avait inventé le photon pour expliquer l'effet photo-électrique; Rutherford, en 1911, avait montré que l'atome était constitué d'un noyau central tout petit contenant la plus grande partie de la masse, les électrons, découverts par Thomson en 1897, devant sans doute graviter autour de ce noyau central.

Quand Bohr vint travailler dans le laboratoire de Rutherford, il vécut dans l'exaltation du nouveau modèle atomique conçu par son maître, mais ne comprenait pas comment il pouvait être dans le vrai. En effet, comment se représenter des électrons qui sont à l'extérieur du noyau? S'ils ne se précipitent pas sur le noyau qui les attire électriquement, il faut admettre qu'ils tournent autour de lui, la force centrifuge contrebalançant la force électrique. « Mais alors, se dit Bohr, qui connaissait à fond les lois de l'électricité de Maxwell, si les électrons bougent, ils doivent émettre des ondes électromagnétiques, perdre sans cesse de leur énergie et, finalement, décrire une spirale et tomber sur le noyau; cette fin brutale doit même se produire en moins de 10^{-8} seconde! Que conclure?

« Par ailleurs, si les électrons sont parfois des émetteurs d'ondes électromagnétiques, en l'occurrence d'ondes lumineuses (quand les atomes sont excités et qu'on les dérange, mais pas dans leur



Émission de la lumière par l'atome

A gauche, **premier temps : excitation de l'atome. L'électron reçoit l'énergie d'excitation qui le fait passer de son orbite circulaire stable à une autre orbite circulaire stable.**

A droite, **deuxième temps : désexcitation. L'électron retombe sur son orbite primitive et reprend son énergie initiale en émettant un photon d'énergie $h\nu$.**

état naturel!), l'électromagnétisme classique voudrait qu'à côté d'une longueur d'onde λ, ils en émettent aussi d'autres $\frac{\lambda}{2}, \frac{\lambda}{3}, ...$

— C'est faux! dit J. J. Balmer et bien d'autres avec lui.

— A ce problème insoluble, moi, Niels Bohr, j'apporte des solutions radicales :

« 1° Je suppose que l'électron décrit autour du noyau une orbite circulaire et que, dans son état normal, il n'émet pas d'ondes électromagnétiques.

« **Certes, je me place en contradiction avec les lois les mieux établies de l'électromagnétisme.** Mais qui me dit que la théorie de Maxwell, valable à l'échelle macroscopique, le soit encore, dans tous les détails, à l'échelle atomique?

« 2° L'émission de la lumière par un atome excité se fait de la façon schématisée ci-dessus, en deux temps.

« Si je suppose qu'il y a émission de lumière par un photon $h\nu$, c'est que je raisonne à l'échelle atomique où la lumière se comporte effectivement comme formée de photons. D'ailleurs, en fait de discontinuité, je suis obligé d'en introduire une autre, si je veux expliquer celles du spectre de raies de l'hydrogène trouvé par Balmer. Je fais donc une troisième hypothèse.

« 3° De toutes les orbites mécaniquement possibles, seules quelques-unes sont réalisées et existent réellement : ce sont

celles où le moment cinétique σ de l'électron (1) est un multiple entier de $\frac{h}{2\pi}$; soit donc $\sigma = \frac{nh}{2\pi}$. Pour la première orbite, on a $n = 1$, pour la seconde $n = 2$, etc.

Ces hypothèses — et surtout la dernière — peuvent paraître bien arbitraires. Mais si je calcule, à partir d'elles, le spectre d'émission de l'atome d'hydrogène, je trouve exactement la formule obtenue expérimentalement par Balmer. Je retrouve la forme $\left(\frac{1}{2^2} - \frac{1}{n^2}\right)$. Je retrouve enfin la même constante R, déterminée cette fois à partir d'autres constantes de la physique, ce qui me permet de rattacher ce phénomène à d'autres, plus généraux. Ce triple accord de l'expérience et de la théorie ne suffit-il pas à justifier les hypothèses faites? »

Cette théorie de Bohr devait vite connaître un grand succès, car, à côté de la série de Balmer, elle en prévoyait également d'autres, qui furent effectivement trouvées, réalisant les prévisions avec une exactitude incroyable. Lyman trouva une deuxième série dans l'ultraviolet, Paschen une troisième dans l'infrarouge. On connaît maintenant deux autres séries et dans la série de Balmer on n'a pas trouvé moins de trente-trois termes.

Ces trois séries de formules similaires —

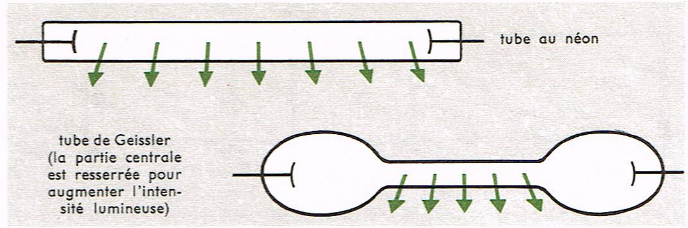
série de Lyman $R\left(\frac{1}{1^2} - \frac{1}{n^2}\right)$; $n = 2, 3, 4, \ldots$;

série de Balmer $R\left(\frac{1}{2^2} - \frac{1}{n^2}\right)$; $n = 3, 4, 5, \ldots$;

série de Paschen $R\left(\frac{1}{3^2} - \frac{1}{n^2}\right)$; $n = 4, 5, 6, \ldots$

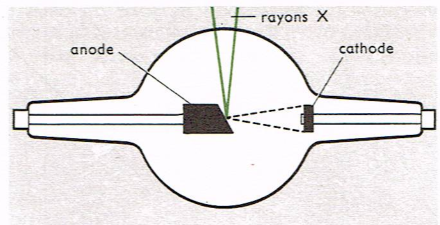
— s'interprètent aisément par le schéma des orbites de Bohr. Se trouvant en si bon chemin, on est tenté de généraliser la théorie de Bohr aux atomes autres que l'hydrogène. Si des modifications et perfectionnements devaient y être apportés, les grandes lignes étaient cependant justes, comme devait le montrer l'étude des rayons X, de leur absorption et de leur émission.

(1) Le moment cinétique de l'électron est la quantité $\sigma = mvr$ (masse × vitesse de l'électron × rayon de l'orbite).



Tubes à émission lumineuse visible. Tension d'alimentation : environ 200 volts; gaz raréfié. Les électrons vont de la cathode à l'anode et excitent sur leur passage les atomes de gaz, qui émettent ensuite de la lumière.

Tube à rayon X à vide très poussé.

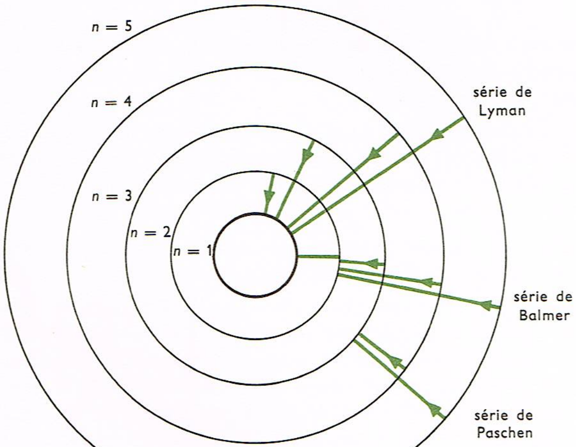


Tension d'alimentation : des dizaines de kilovolts; vide poussé. Les électrons fortement accélérés vont se précipiter avec violence sur l'anode, qui émet les rayons X en retour.

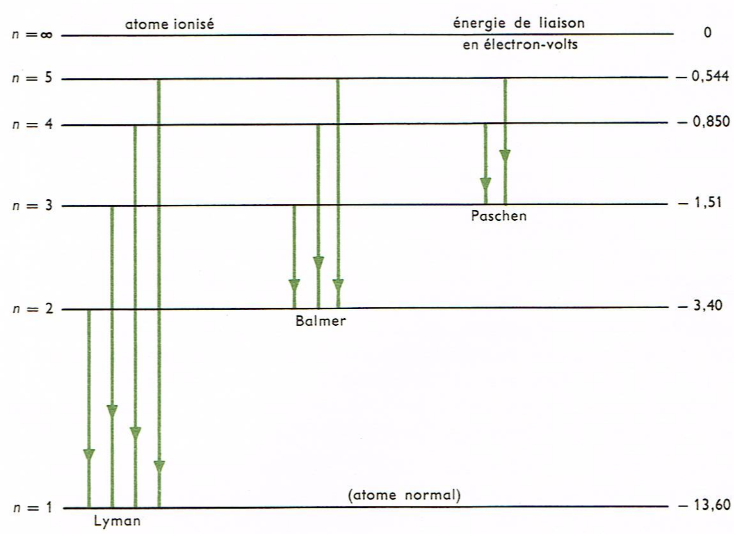
Les électrons doivent se distribuer sur plusieurs orbites circulaires de Bohr, ou, disons, sur plusieurs couches (nous verrons plus loin combien il peut y en avoir par couche), ceux des couches internes étant évidemment plus liés au noyau central que ceux des couches externes. L'absorption des rayons X par la matière, dont nous avons déjà parlé plus haut, nous avait fait parler de niveaux d'énergie des électrons de l'atome. Ces niveaux, appelés alors K, L, M, correspondent tout simplement aux différentes couches de Bohr (la couche K correspondant à $n = 1$, etc.).

D'un autre côté, plus l'atome est lourd, plus les orbites circulaires doivent être proches du noyau central, l'augmentation des charges électriques positives de ce dernier devant augmenter l'attraction électrostatique sur les électrons. Ainsi

Représentation des orbites de Bohr.



Représentation des orbites de Bohr en niveaux d'énergie; chaque orbite correspond à un niveau d'énergie, les couches internes étant les plus liées au noyau et donc les plus stables.



les mêmes couches correspondent à des niveaux d'énergie plus grands dans les atomes lourds que dans l'atome d'hydrogène. On comprend alors pourquoi il faut des rayons X pour mettre en évidence les couches internes, la lumière visible étant tout juste capable d'exciter les couches externes. Dans un même ordre d'idées, il faut des moyens d'excitation plus importants pour produire des rayons X que des rayons lumineux visibles (heureusement d'ailleurs), et aussi un tube plus gros qu'un simple tube au néon.

Que se passe-t-il dans une telle émission? Les électrons fortement accélérés arrivent à pénétrer dans les couches internes de l'atome et à en éjecter des électrons. Il se produit alors un réarrangement électronique, un saut confus d'électrons de couche en couche et, par le fait même, une émission de rayons X, dont l'énergie correspond à celle des couches internes. Ainsi, plus lourd est l'atome frappé, plus énergique est le rayonnement émis.

l'atome de Sommerfeld

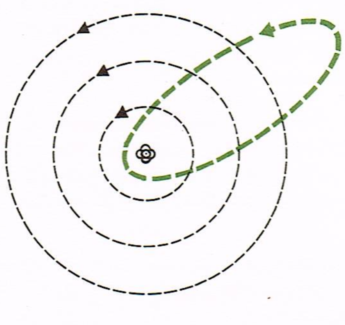
Mais revenons à nos chers spectroscopistes, qui furent sans s'en douter à l'origine de tous ces développements nouveaux de la physique. A la suite de Balmer, et peu de temps après, ils essayèrent de systématiser comme lui les spectres d'éléments autres que l'hydrogène; ils y arrivèrent pour les métaux alcalins (sodium, potassium), qui présentent certaines analogies chimiques avec l'hydrogène et se trouvent dans la même colonne de la classification périodique. Les formules trouvées ressemblaient à celle de Balmer, mais avec des corrections supplémentaires, et là où Balmer ne voyait qu'une série, ils en observaient plusieurs. La théorie de Bohr n'expliqua pas cet état de choses, de même qu'elle n'expliqua pas la subdivision des niveaux d'énergie L, M en « sous-niveaux » dans les expériences avec rayons X.

Ce fut Sommerfeld qui en donna une explication plausible. Bohr fait de l'atome un système solaire en miniature; de l'électron, un satellite du noyau; mais alors pourquoi, à l'égal de notre Terre, cette petite terre n'aurait-elle pas autour de son soleil une orbite elliptique? L'orbite ciculaire de Bohr ne serait alors qu'un cas particulier d'une orbite elliptique plus générale, que Sommerfeld caractérisa par deux nombres :

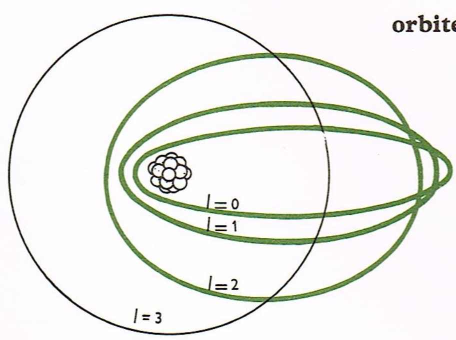
— le nombre *n*, le même que celui introduit par Bohr, et qui mesure le grand axe de l'ellipse (ou le diamètre du cercle);

— un nombre *l*, qui en fait varier la forme, plus ou moins allongée.

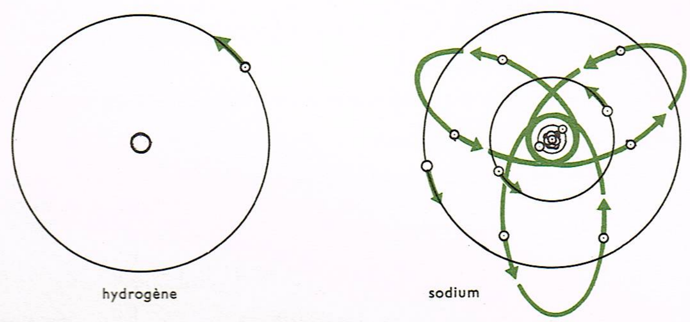
Sommerfeld montra que dans l'hydrogène, où il y a un seul électron, l'énergie de cet électron, et donc la raie d'émission, ne dépendent pas de la forme de l'ellipse, mais seulement de la longueur de son grand axe et donc seulement du nombre *n* de la théorie de Bohr, qui se trouve ainsi être rigoureuse. Dans un atome plus compliqué, ce n'est plus vrai, car un électron qui a une orbite elliptique allongée pénètre dans les couches électroniques internes et son énergie doit être perturbée par leurs électrons.



orbites de Sommerfeld pour $n = 4$



Représentation d'atomes en orbites de Sommerfeld.



Bohr avait quantifié le nombre *n* pour expliquer les raies de l'hydrogène, Sommerfeld, à son tour, quantifie le nombre *l*, lui imposant d'être entier et de ne pouvoir prendre que les valeurs 0, 1, 2, ..., $n - 1$ (il ressort de sa théorie que *l* ne peut pas être plus grande que $n - 1$), moyennant quoi il retrouve les raies plus compliquées des métaux alcalins.

Des développements ultérieurs devaient montrer que c'était le nombre *l*, et non *n*, qui mesurait le moment cinétique de l'électron dans sa révolution autour du noyau, moment cinétique qu'on représente souvent par un vecteur perpendiculaire au plan de l'orbite.

le spin de l'électron

Ce schéma pourtant ne suffisait pas encore à tout expliquer; une analyse spectroscopique plus fine montrait, en effet, que chaque raie du spectre des alcalins était en réalité double.

Qui ne connaît, pour avoir roulé de nuit sur certains tronçons de nos routes, les lampes à sodium et leur lumière jaunâtre? Eh bien, la raie jaune caractéristique qu'elles émettent est, en fait, double : $\lambda = 5890$ A et $\lambda = 5896$ A ($1 A = 10^{-8}$ cm).

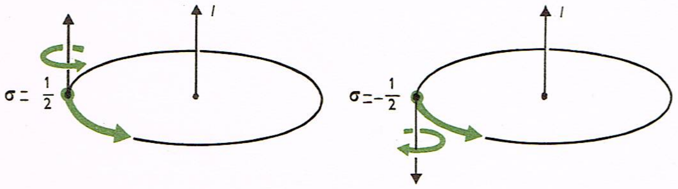
D'ailleurs, l'absorption des rayons X, encore elle, montre qu'il doit y avoir un niveau dans la couche K et trois dans la couche L; la théorie de Sommerfeld donne bien un niveau dans la couche K, mais deux seulement dans la couche L :

couche K $n = 1$ $l = 0$
couche L $n = 2$ $l = 0$
$l = 1$.

Il fallut encore se torturer l'esprit pour chercher une autre raison à ces « anomalies ». Une hypothèse hardie fut avancée en 1925 par les Scandinaves Uhlenbeck et Goudsmit. Au lieu de voir l'électron comme un simple point matériel, ceux-ci en font une petite boule qui tourne sur elle-même, en même temps qu'elle tourne autour du noyau (l'analogie avec la Terre devient presque parfaite!). Un tel électron doit alors posséder un moment cinétique propre, qu'ils appellent « spin » (du verbe

anglais *to spin*, tourner), et auquel ils donnent la valeur $\frac{1}{2}$ (1).

Ce moment cinétique propre $\frac{1}{2}$ doit se combiner au moment cinétique l dû à la révolution de l'électron autour du noyau; Uhlenbeck et Goudsmit supposent que le moment cinétique total j qui en résulte peut seulement prendre les valeurs $l + \frac{1}{2}$



Les deux seules orientations du spin d'après Uhlenbeck et Goudsmit.

ou $l - \frac{1}{2}$, comme si les deux rotations se faisaient dans le même sens ou dans le sens opposé.

Moyennant ces hypothèses, ils expliquent les doublets des raies; ils expliquent également les trois niveaux de la couche L comme on peut le voir facilement :

$$\text{couche K} \quad n = 1 \quad l = 0 \quad j = \frac{1}{2};$$

$$\text{couche L} \quad n = 2 \quad \begin{cases} l = 0 & j = \frac{1}{2}; \\ l = 1 & \begin{cases} j = \frac{1}{2}; \\ j = \frac{3}{2}. \end{cases} \end{cases}$$

Cette hypothèse du spin fut brillamment confirmée par Stern et Gerlach dans une expérience qui apporta également de nouveaux renseignements sur les lois bizarres de ce petit monde.

Si l'électron tourne autour du noyau, il provoque un petit courant, et l'atome devient une bobine électrique en miniature. Tout comme une bobine électrique ou une aiguille aimantée, il doit subir l'influence d'un champ magnétique extérieur. Si l'on réussit à mettre en évidence cette influence, on peut, par le fait même, mesurer le « moment magnétique » de l'atome, qui doit être proportionnel, comme il est facile de le montrer, à son moment cinétique total.

La spectroscopie indique par ailleurs qu'il existe certains atomes ayant un seul électron sur la couche externe, dont le moment cinétique total j doit se réduire au spin, s'il existe, de cet électron externe. L'atome d'argent est de ceux-là, et pourrait servir ainsi à mesurer le spin de l'électron.

En partant de telles considérations, Stern et Gerlach mirent au point une expérience pour mettre en évidence ce fameux spin. Une boule d'argent qu'ils chauffaient à très haut degré émettait, dans le vide, des atomes d'argent dans toutes les directions; un faisceau fin de tels atomes, délimité par deux fentes F_1 et F_2, passait dans l'entrefer d'un électro-aimant, qu'ils avaient spécialement construit très dissymétrique pour augmenter l'effet possible. Sans courant dans l'électro-aimant, comme on pouvait s'y attendre, l'argent venait se déposer, juste dans la direction des deux fentes en O, où il formait une tache. En présence d'un courant, ce dépôt disparaissait, mais était remplacé par deux nouvelles taches très petites toujours, situées de part et d'autre de O et à la même distance. Sans aucun doute, il fallait admettre que l'atome d'argent possédait effectivement un moment magnétique et l'électron un spin, mais l'existence de ces deux taches n'était-elle pas incompréhensible? Si spin il y avait, il devrait sans doute être orienté au hasard, et comme l'influence du champ magnétique dépend de cette orientation, il devrait y avoir une tache diffuse étalée autour de O, non ces deux taches discontinues! Que se passe-t-il?

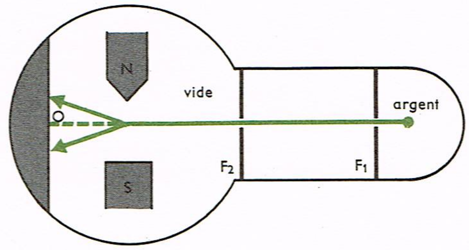


Schéma de l'expérience de Stern-Gerlach.

Si je suppose le spin de l'électron orienté dans le sens du champ magnétique, ou dans le sens opposé, la mesure de la déviation produite me donne toujours la valeur $\frac{1}{2}$ pour ce spin, comme prévu par Uhlenbeck et Goudsmit.

Mais pourquoi ce spin peut-il se trouver orienté dans deux directions différentes? En fin de compte, m'habituant aux discontinuités de ce monde atomique, ne puis-je pas timidement me risquer à dire, d'un langage certes un peu énigmatique : « La composante du spin suivant le champ magnétique se trouve quantifiée et ne peut prendre que les valeurs $+\frac{1}{2}$ et $-\frac{1}{2}$ »?

C'est la réflexion que firent les physiciens de l'époque, aiguillés d'ailleurs par le fait que cette hypothèse les aidait à comprendre un phénomène de décomposition spectrale par le champ magnétique, le phénomène connu depuis 1896 sous le nom d'« effet Zeeman ». Là aussi les raies se décomposaient, non en taches diffuses, mais en d'autres raies plus nombreuses, mais tout aussi fines.

Toutes ces propriétés furent parfaitement comprises quand on admit la loi générale suivante, qui peut paraître bien étrange à première vue :

« Un moment cinétique j, quantifié, admet, suivant une direction quelconque (par exemple, celle du champ magnétique), une composante m, qui, elle aussi, est quantifiée, selon la règle qui n'admet pour m que les valeurs

$$-j, -j+1, -j+2, \ldots, j-2, j-1, j. \text{ »}$$

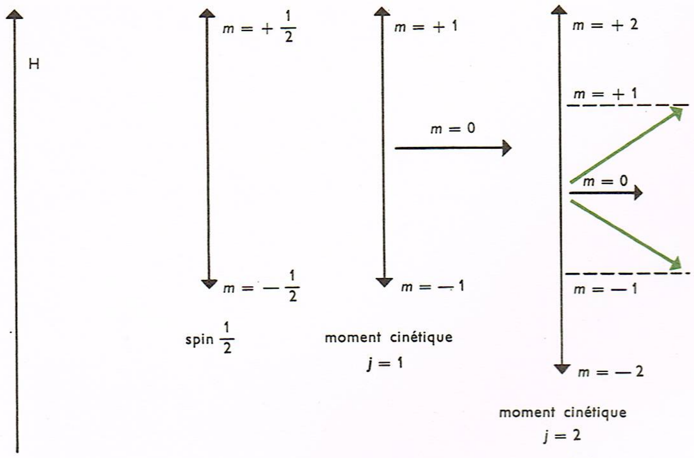
Tout se passe comme si, sous l'influence du champ magnétique, le moment cinétique rectifiait sa position pour choisir entre les seules qui lui sont permises.

A y regarder de près, cette règle de quantification n'est pas plus bizarre que les autres déjà rencontrées, mais bien du même type. **Le mouvement de l'électron dans l'atome n'est pas complètement libre, mais il est astreint à certaines conditions qui n'ont pas leur équivalence dans le monde qui tombe immédiatement sous nos yeux.** Ces conditions peuvent s'exprimer par des discontinuités qui recouvrent celles des nombres entiers.

Ainsi, au terme de ces perfectionnements de détail successifs apportés à l'explication de la répartition des raies dans les spectres atomiques, il se trouve qu'un électron qui tourne autour du noyau atomique peut être décrit par quatre nombres quantiques :

1° Le nombre n de Bohr, qui numérote la couche et donne l'ordre de grandeur du niveau d'énergie : $n = 1, 2, \ldots$;

(1) Plus exactement la valeur $\frac{1}{2} \times \frac{h}{2\pi}$, que par abréviation on appelle simplement $\frac{1}{2}$, de même que $l = 1$ signifie en réalité $l = 1 \times \frac{h}{2\pi}$.



Orientations possibles d'un moment cinétique suivant la direction du champ magnétique H.

2° Le nombre l, qui mesure le moment cinétique de l'électron dans sa révolution autour du noyau : $l = 0, 1, 2, ..., n - 1$;

3° Le nombre j, qui mesure le moment cinétique total de l'électron, combinaison du nombre l et du nombre $\frac{1}{2}$, qui mesure le spin, c'est-à-dire le moment cinétique de l'électron dans sa révolution sur lui-même : $j = l + \frac{1}{2}$ ou $j = l - \frac{1}{2}$.

4° Le nombre m, qui mesure la composante du moment cinétique total j suivant un axe de référence. Il ne peut prendre que les $(2j + 1)$ valeurs suivantes :

$$m = -j, -j + 1, ..., j - 1, j.$$

Il se trouve qu'en supposant qu'un seul électron puisse avoir une série de valeurs données pour ces nombres n, l, j, m, on peut prévoir comment se bâtissent les atomes et — ce qui viendra confirmer la théorie tout entière — on retrouve et on explique les mystérieuses régularités rencontrées dans la classification périodique de Mendéléiev.

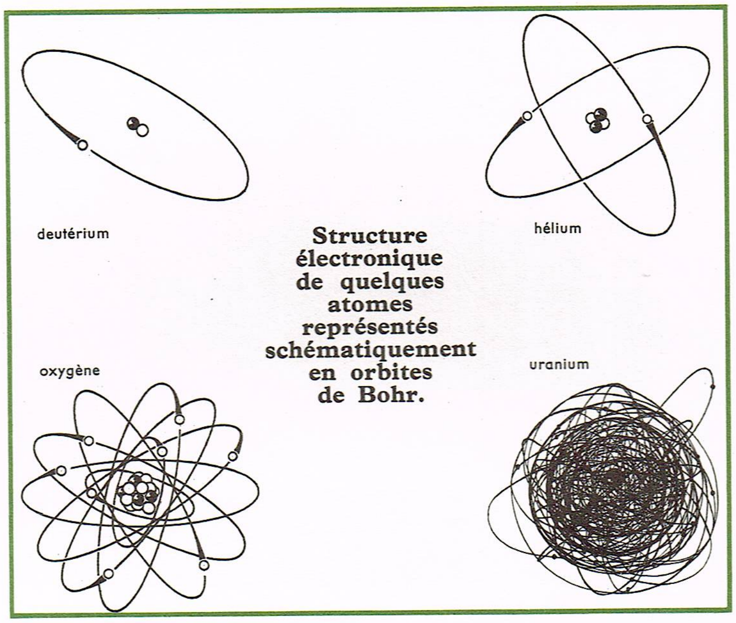
l'édification électronique des atomes

Reconstruisons maintenant, pierre par pierre, l'édifice électronique des atomes. Les électrons occupent d'abord les couches les plus stables, celles où ils sont unis plus intimement au noyau; ils remplissent ainsi les couches de Bohr successives, en commençant par la première.

La première se remplit donc avec deux électrons, ce qui coïncide avec la première période de la classification périodique, limitée aux seuls hydrogène et hélium. Puis vient une deuxième période, correspondant au remplissage progressif de la deuxième couche, qui se trouve complète avec le néon. La troisième couche commence alors à s'édifier, et comprend à la fin de la troisième période huit électrons avec l'argon. Le remplissage se continue ensuite, mais de façon un peu plus compliquée, car les électrons préfèrent de temps en temps édifier une nouvelle couche au lieu de terminer celle qui était en cours. C'est ainsi que si les quatorze métaux dits des « terres rares » (métaux qui se trouvent en quantité assez faible dans certaines terres) ont des propriétés chimiques très voisines, ils le doivent au fait que leurs deux couches externes sont identiques, et dans le dernier de ces métaux une couche précédente se complète finalement avec les quatorze électrons qui manquaient par exemple au xénon.

Cette édification continuelle de nouvelles couches vient expliquer la répétition des propriétés chimiques dont la classification périodique se voulait être le reflet. Si les gaz rares de l'air (hélium, néon, argon, crypton, xénon) forment une famille d'éléments très peu réactifs, c'est qu'ils possèdent tous huit électrons sur leur couche externe, ce qui leur confère une grande stabilité. Le sodium et les métaux alcalins, au contraire, doivent leur grande réactivité à l'électron isolé qui — pourrait-on dire — se meurt d'ennui sur sa couche externe. C'est ce seul électron supplémentaire, ô ironie, qui rend si différents, tant du point de vue physique que chimique, deux éléments comme le néon et le sodium!

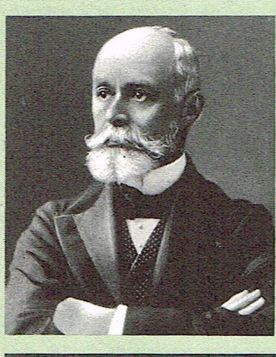
La chimie ne devient plus qu'un jeu, complexe d'ailleurs, d'électrons et de liaisons électroniques, complexe, on le comprend déjà par le petit aperçu des propriétés que nous venons de voir. Encore n'avons-nous pas compris leur raison d'être, imprévisibles qu'elles sont par les méthodes de la physique classique de Newton et de Maxwell. Il faudra inventer à l'usage des électrons une nouvelle mécanique (qu'on appellera « quantique », on comprend maintenant pourquoi), qui substituera à un monde continu un univers discontinu, à l'image de l'atome lui-même.



Structure électronique de quelques atomes représentés schématiquement en orbites de Bohr.

				nombre de composantes m	nombre d'électrons par couche
1re couche	$n = 1$	$l = 0$	$j = \frac{1}{2}$	2	2
2e couche	$n = 2$	$l = 0$	$j = \frac{1}{2}$	2	8
		$l = 1$	$j = \frac{1}{2}$	2	
			$j = \frac{3}{2}$	4	
3e couche	$n = 3$	$l = 0$	$j = \frac{1}{2}$	2	18
		$l = 1$	$j = \frac{1}{2}$	2	
			$j = \frac{3}{2}$	4	
		$l = 2$	$j = \frac{3}{2}$	4	
			$j = \frac{5}{2}$	6	

etc.

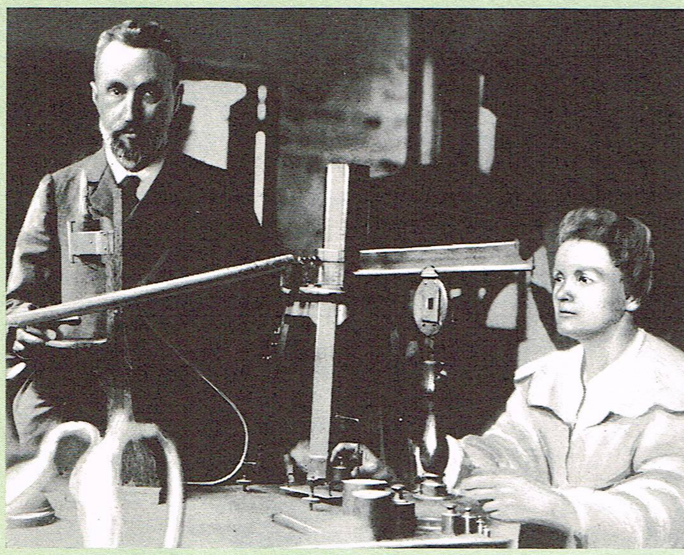


Henri Becquerel.

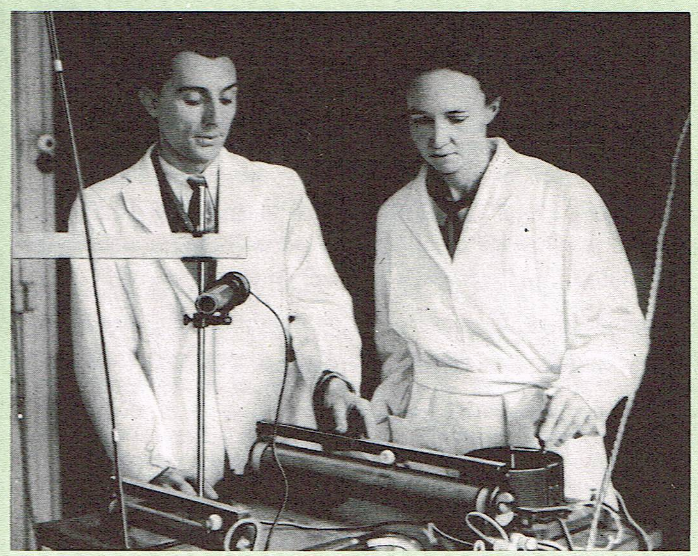


« Je me souviens du ravissement que nous éprouvions lorsqu'il nous arrivait d'entrer la nuit dans notre domaine et que nous apercevions de tous les côtés les silhouettes faiblement lumineuses de notre travail. »

Marie Curie (au 25e anniversaire de la découverte du radium).



Pierre et Marie Curie dans leur laboratoire.



Frédéric et Irène Joliot-Curie.

la physique nucléaire

la radio-activité naturelle

Becquerel, P. et M. Curie, les premiers pionniers

Revenant encore à notre édification électronique, remarquons que le cortège d'électrons, loin de s'étendre à l'infini, ne dépasse guère la centaine; pourquoi ne connaissons-nous guère plus d'une centaine d'atomes différents? La réponse n'est pas déterminée par les électrons. C'est au cœur de l'atome qu'il faut la chercher : le nombre de charges positives du noyau doit évidemment augmenter de pair avec le nombre d'électrons, et alors le noyau devient si peuplé — crise de logement en quelque sorte — qu'il en éclate de toutes parts.

Ici s'ouvre un nouveau chapitre de la physique, celui de la physique nucléaire, qui mettra en évidence ces immenses réserves d'énergie du noyau atomique.

L'histoire de la physique nucléaire commença de façon accidentelle et inattendue, presque cocasse, avec la découverte de la radio-activité naturelle.

En faisant des expériences sur les décharges électriques à haute tension dans les tubes à vide, Röntgen venait de découvrir, en 1895, des radiations nouvelles capables de traverser des corps opaques à la lumière. Dans son mémoire, il notait en particulier que ces rayons X, bien mystérieux, semblaient provenir de la paroi du tube, qui était phosphorescente. (En fait, Röntgen travaillait avec un tube à vide sans anode spéciale, et les électrons venaient souvent frapper la paroi du tube, produisant à la fois la phosphorescence du tube et les rayons X, mais il ne pouvait guère s'en douter à une époque où les électrons n'étaient même pas encore découverts.) En France, Henri Becquerel, spécialiste des problèmes de luminescence, et Henri Poincaré, homme universel, s'intéressèrent tout de suite à la découverte et crurent que les rayons X étaient dus à la phosphorescence du tube. Becquerel voulut vérifier cette hypothèse avec l'arsenal de matières fluorescentes qu'il avait accumulées au cours de ses travaux. Il se trouva qu'il

Phot. Délius, Keystone.

avait parmi elles un échantillon de minerai d'uranium, la pechblende de Bohême, légèrement fluorescent.

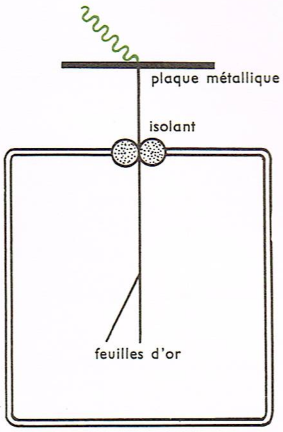
Eclairé par une forte lumière, puis enveloppé d'un papier noir, ce corps devenait effectivement capable d'impressionner une plaque photographique. Il crut avoir abouti dans ses recherches quand, à quelque temps de là, voulant employer d'autres plaques photographiques vierges, il s'aperçut qu'elles étaient voilées. Au lieu d'incriminer le fabricant et d'en rester là, il fut frappé par le fait que ces plaques avaient traîné longtemps dans un tiroir au voisinage d'un autre échantillon de minerai d'uranium qui n'avait pas encore servi. Il venait ainsi de découvrir que son échantillon était capable d'émettre de lui-même, sans influence extérieure, des radiations, donc qu'il était radio-actif (producteur de rayons).

Très rapidement, il observa les premières propriétés de ce rayonnement, qu'il croyait d'abord constitué de rayons X : celui-ci traversait les corps opaques, faisait dévier les feuilles d'or d'un électroscope, rendait l'air conducteur d'électricité. Quel était donc ce nouveau rayonnement si énergique?

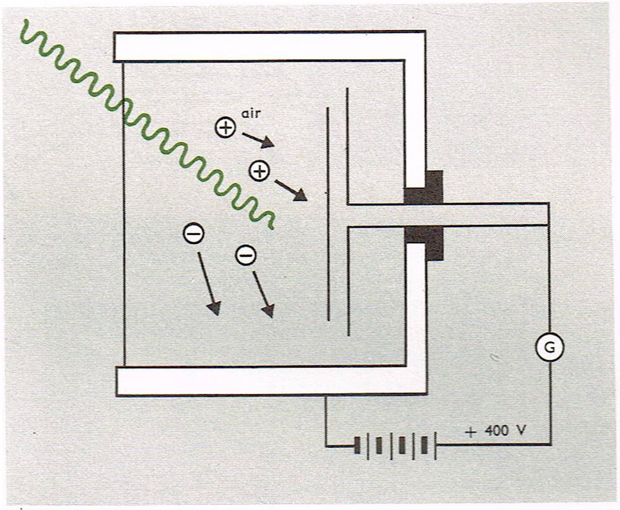
Pendant ce temps, Marie Sklodowska, future Mme Curie, encore jeune étudiante, était à la recherche d'un sujet de travail utilisable pour sa thèse de doctorat. Quand elle apprit la découverte de Becquerel, elle décida de se lancer dans ce nouveau domaine. Elle commença par faire une recherche et une étude systématiques des corps chimiques qui pouvaient posséder une si mystérieuse propriété. Pour pouvoir mieux détecter et mesurer le rayonnement émis, elle utilisa une chambre d'ionisation, beaucoup plus sensible qu'un simple électroscope. Elle trouva peu de corps radio-actifs, mais tous les composés d'uranium et également tous les composés de thorium l'étaient. L'étrange rayonnement restait, par ailleurs, insensible aux traitements physiques ou chimiques les plus énergiques que l'on pouvait imaginer, et devait certainement avoir son origine au plus profond secret de l'atome d'uranium ou de thorium. Cependant, se situant dans cette perspective, Mme Curie fut intriguée par certains résultats : à poids égal d'uranium, certains minerais étaient plus actifs que d'autres. La pechblende en particulier (un oxyde d'uranium) faisait montre d'une activité remarquable, et Marie Curie pensa que peut-être ce minerai contenait un élément, inconnu jusqu'alors, plus actif que l'uranium lui-même. Elle et son mari, Pierre Curie, se mirent alors à l'œuvre pour essayer de mettre en évidence ce nouvel élément : entreprise difficile, car la précision des analyses chimiques déjà faites sur la pechblende leur permettait d'espérer une proportion en poids de l'élément recherché égale au plus à 1 p. 100.

Deux ans furent nécessaires pour séparer les différents corps chimiques de la pechblende et pour essayer d'isoler, de fractionnement en fractionnement, des quantités de matière de plus en plus radio-actives. L'élément recherché se trouvait, en fait, être dans une proportion plus petite que celle qui était espérée, mais ce travail acharné et méticuleux eut finalement sa récompense. Sans les avoir isolés, les Curie annoncèrent la découverte de deux nouveaux éléments radio-actifs. Le premier fut trouvé dans la fraction contenant le bismuth; ils lui donnèrent le nom de « polonium », en l'honneur du pays d'origine de Marie Curie. Le second, trouvé dans la fraction contenant le baryum, était encore plus radio-actif que le précédent, et ils l'appelèrent le « radium ». On se rendra compte de l'œuvre gigantesque entreprise et menée à bien par eux, quand on saura que, commencés en 1896, leurs travaux les amenèrent en 1902, après traitement de 2 tonnes de minerai d'uranium, à isoler quelques milligrammes de chlorure de radium pur. La mesure du poids atomique du radium leur permit alors de le mettre à sa place dans la classification périodique.

Par ces travaux, une méthode chimique était mise au point et l'élan était donné. Dès 1904, comprenant vite l'intérêt médical qui pouvait en être tiré, on commença à isoler industriellement le radium. Les recherches devaient s'en trouver considérablement raccourcies.



Electroscope : le rayonnement radio-actif tombe sur la plaque métallique et la charge électriquement soit parce qu'il transporte des charges électriques, soit parce qu'il arrache des électrons au métal. Les feuilles d'or très légères, qui sont au contact de la plaque, se séparent par répulsion électrostatique.

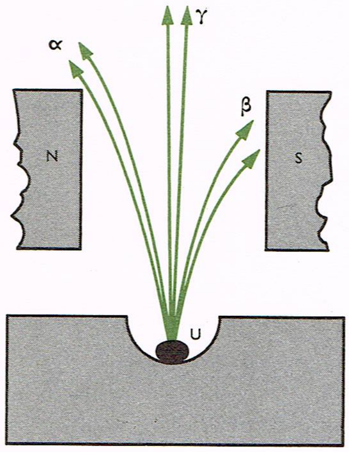


Chambre d'ionisation : si une particule traverse la chambre d'ionisation, elle ionise sur son passage les molécules d'air : les ions positifs et négatifs formés vont respectivement aux électrodes négative et positive; le courant produit peut être mesuré par un galvanomètre sensible. Le rayonnement radio-actif est si énergique que, dans des conditions habituelles, une particule peut créer de 100 à 20 000 ions!

la nature du rayonnement radio-actif

En même temps que, à l'exemple de Pierre et de Marie Curie, on cherchait, par des moyens chimiques et spectrographiques, à isoler et à identifier de nouveaux radio-éléments, ainsi que les éléments qui naissaient des transformations radio-actives, on essayait de percer le mystère du rayonnement émis, mais un rayonnement aussi énergique n'avait jamais été observé, et les progrès n'étaient vraiment possibles que si l'on pouvait compter et localiser les particules émises. **On devait assister ainsi par la force des choses à un gros effort dans la recherche et la mise au point de techniques nouvelles.**

On s'aperçut certes rapidement que le rayonnement d'un minerai radio-actif était complexe, puisque, sous l'effet d'un champ magnétique, il se divisait en trois parties. Un premier rayonnement, dit α, dévié dans un certain sens, devait être constitué

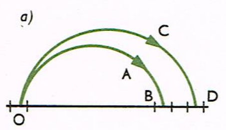


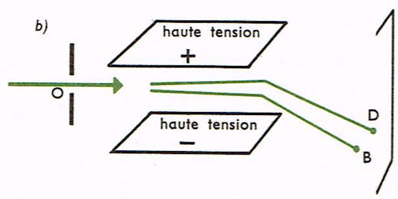
Effet d'un champ magnétique perpendiculaire au plan de la figure sur le rayonnement émis par l'uranium U.

Rayons α = atomes d'hélium ayant perdu leurs deux électrons (α = He^{++}).

Rayons β = électrons.

Rayons γ = analogues aux rayons X, mais de longueur d'onde plus petite.





Action du champ magnétique. Action du champ électrique.
La particule la plus lourde est la moins déviée.

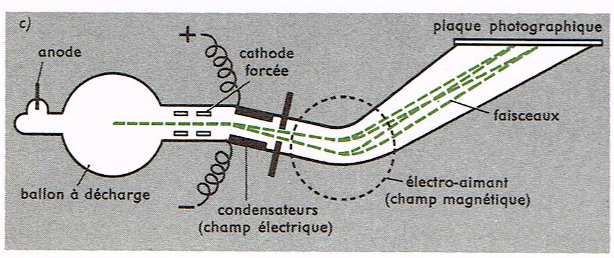


Schéma d'un spectrographe de masse.

a) **Sous l'effet du champ magnétique, les particules décrivent des arcs de cercle de rayon d'autant plus grand que la particule est plus lourde.**

b) **Sous l'effet du champ électrique régnant sur les deux plateaux du condensateur, les particules décrivent les mêmes courbes (paraboles) que les corps pesants sous l'effet de la pesanteur. Mais l'analogie s'arrête là : la particule est d'autant moins déviée qu'elle est plus lourde.**

c) **Les atomes, ionisés dans le ballon à décharge, sont attirés par la cathode, percée, qu'ils traversent. Ils sont ensuite déviés par les champs électrique et magnétique et vont frapper la plaque photographique d'autant plus loin que leur masse est plus grande.**

de charges électriques positives; un deuxième, dit β, dévié dans un sens opposé, de charges électriques négatives. Le troisième, dit γ, non dévié, devait venir de particules neutres, bien qu'on se fût demandé d'abord s'il ne s'agissait pas de particules chargées trop rapides pour être courbées.

La nature du rayonnement β fut même assez vite connue, en répétant la méthode qui avait permis à J. J. Thomson, en 1897, de découvrir l'électron dans le courant électrique qui traversait les gaz; en utilisant les actions croisées d'un champ magnétique et d'un champ électrique, il est possible, en effet, de mesurer en même temps la vitesse et le rapport $\frac{e}{m}$ de la particule. On trouva la même valeur $\frac{e}{m}$ que Thomson, mais les électrons avaient une vitesse et donc une énergie beaucoup plus grandes, de l'ordre de 1 million d'électron-volts (1 MeV). Il est bon ici d'avoir des points de comparaison : un tube producteur de rayons X fonctionne couramment sous une tension de 50 000 V, et l'énergie des électrons accélérés dans ce tube est ainsi de 50 000 eV. L'énergie de liaison des électrons les plus externes de l'atome, et partant l'énergie mise en jeu par atome, dans les réactions chimiques, est de 1 à 10 eV. Il est vrai que le nombre d'atomes intervenant dans une réaction chimique est beaucoup plus important que dans la radio-activité naturelle. La différence d'échelle n'en est pas moins frappante.

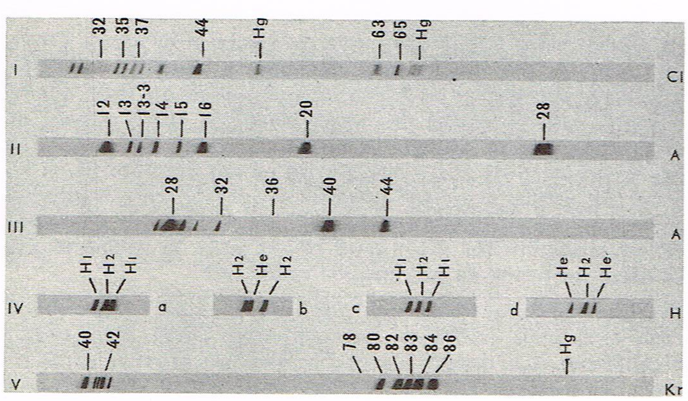
La nature des rayons α, moins déviés que les rayons β, fut de ce fait plus difficile à établir. Il fallut faire appel à des champs électromagnétiques plus intenses, mais la précision obtenue dans la mesure de $\frac{e}{m}$ ne fut pas alors suffisante pour décider s'il s'agissait du noyau d'hélium He ou de celui d'hydrogène H. Il fallut donc déterminer également si la particule contenait une charge double (He^{++}) ou simple (H^{+}). Ce problème, qui obligeait à accumuler les particules en grand nombre pour mesurer leur charge totale, ne fut donc possible que quand on sut les compter une à une.

La nature du rayonnement γ semblait être la même que celle de rayons X très durs, tellement les propriétés se ressemblaient. Ce n'est pourtant qu'à partir de 1913 que l'on put montrer (par la découverte de la diffraction sur les cristaux) qu'on avait affaire, dans un cas comme dans l'autre, à une onde électromagnétique, la longueur d'onde des rayons γ étant simplement plus petite que celle des rayons X.

quelques techniques de mesure

les spectrographes de particules

A la fin du XIXe siècle, la découverte des électrons, celle des rayons X et celle de la radio-activité naturelle devaient être l'occasion de l'élaboration de nouvelles techniques propres à l'étude de particules se déplaçant dans l'air ou dans le vide. Le spectrographe de masse, inventé par J. J. Thomson, lui avait fait découvrir les électrons et lui avait indiqué qu'un même élément chimique pouvait contenir deux éléments de masses différentes; les isotopes (littéralement « même place », sous-entendu « dans la classification périodique ») furent ainsi « vus » pour la première fois. Les perfectionnements apportés à la méthode consistèrent à mettre au point des appareils



Spectres de masse du chlore (partiel), de l'argon (A), de l'hydrogène (H) et du krypton (Kr).

donnant pour un ion de masse donnée une image très fine sur la plaque photographique et à améliorer le pouvoir séparateur de l'instrument. Plus simple, mais très utile aussi, fut également la mise au point du spectrographe de vitesse destiné à analyser les vitesses de particules de masse identique et connue. Les spectres obtenus ressemblent aux spectres optiques déjà rencontrés.
L' « optique électronique » était née.

compteurs à scintillations

Ce fut à propos des rayons α que, pour la première fois, on compta les particules une à une. Pierre et Marie Curie avaient employé une chambre d'ionisation, qui mesurait l'intensité radio-active totale. Le rayonnement pouvait se manifester également soit par impression photographique, soit par production de fluorescence. C'était là pourtant plus un moyen de détection que de mesure. Cependant, en examinant de près une couche de sulfure de zinc rendue phosphorescente par bombardement de rayons α, on s'aperçut que de petites étoiles lumineuses naissaient puis disparaissaient : la surface sensible était couverte de scintillations, marques d'impact de rayons α; on pouvait voir en quelque sorte et compter les particules du rayonnement : ce fut cette méthode qu'utilisa en particulier Rutherford pour mesurer l'intensité du rayonnement α après diffusion à travers une feuille d'or, et qui lui valut en 1911 la découverte du noyau atomique. Une telle méthode obligeait évidemment à employer de faibles intensités pour pouvoir compter les impacts avec succès; l'œil se fatiguait vite également. Plus tard, on remplaça l'œil par un photomultiplicateur, et le compteur à scintillations devint un instrument automatique.

compteur de Geiger

Dans le même temps, et toujours en Angleterre, sous l'impulsion de Rutherford, Geiger mit au point un appareil, dérivé un peu de la chambre à ionisation, mais qui pouvait compter les particules une à une. Le tube est porté à ce qui est appelé un « potentiel disruptif », c'est-à-dire qu'il suffit du passage d'une particule pour amorcer l'ionisation de l'air, et la décharge électrique éclate entre le fil métallique central et l'enveloppe extérieure du tube. Il y a alors passage d'un courant dans le circuit extérieur, où Geiger intercalait une forte résistance. Le rôle de cette résistance est facile à comprendre : traversée par le courant, elle provoque dans le fil central une brusque chute de potentiel, qui arrête la décharge. Le tube est alors prêt à fonctionner de nouveau et à enregistrer le passage d'une autre particule. On a pu augmenter son taux de répétition au prix d'une complication des circuits électroniques. De même, on a pu le rendre plus ou moins sensible à diverses radiations en le remplissant avec différents gaz. En pratique, sa souplesse d'emploi en a fait un instrument précieux de détection et de mesure aussi bien du rayonnement radio-actif que du rayonnement cosmique.

chambre de Wilson (1911)

L'étude de la radio-activité naturelle avait déjà bien progressé, quand, en 1911, Wilson montra qu'on pouvait, avec son appareil, « voir » en quelque sorte les trajectoires des rayons. La chambre de Wilson devait vite connaître un grand succès : la possibilité de voir les traces des particules et de suivre leur évolution constituait un avantage essentiel sur tout autre appareil de détection.

Le principe de fonctionnement repose sur le fait que les particules électrisées ionisent les molécules d'air sur leur passage et créent des ions. Si l'on envoie ces particules dans un récipient (la chambre de Wilson) contenant de l'air sursaturé de vapeur d'eau, et que l'on fasse brusquement une détente qui refroidisse le gaz, les gouttelettes d'eau se forment d'abord

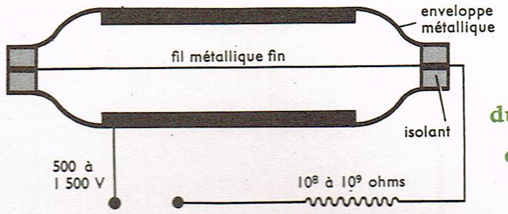
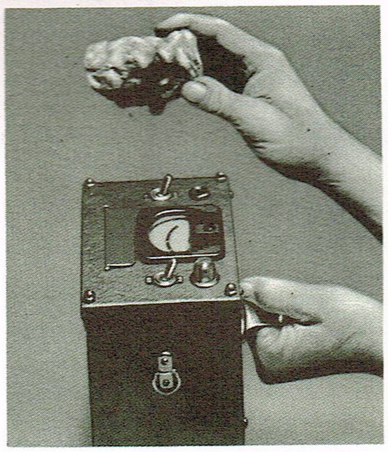
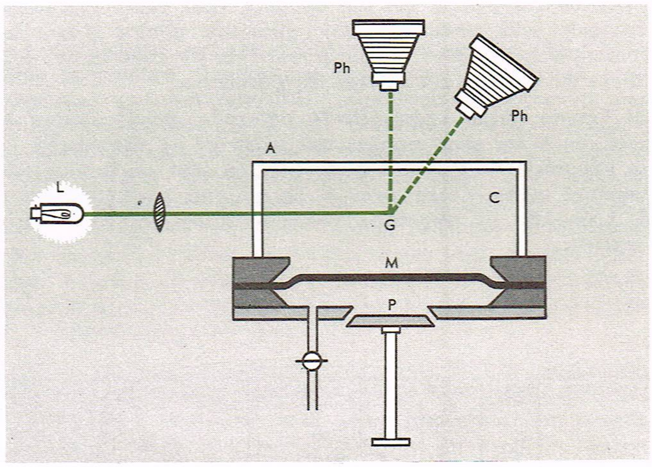


Schéma du compteur de Geiger.

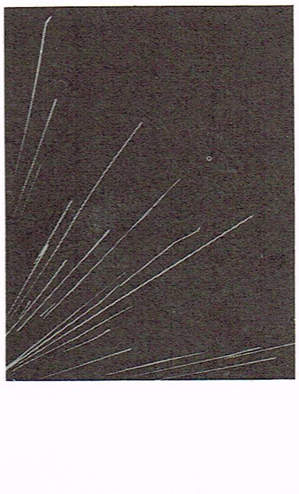


En approchant du compteur de Geiger un corps radio-actif (minerai d'uranium, par exemple), l'appareil crépite. Compteur et dispositif d'amplification sont ici enfermés dans un boîtier.



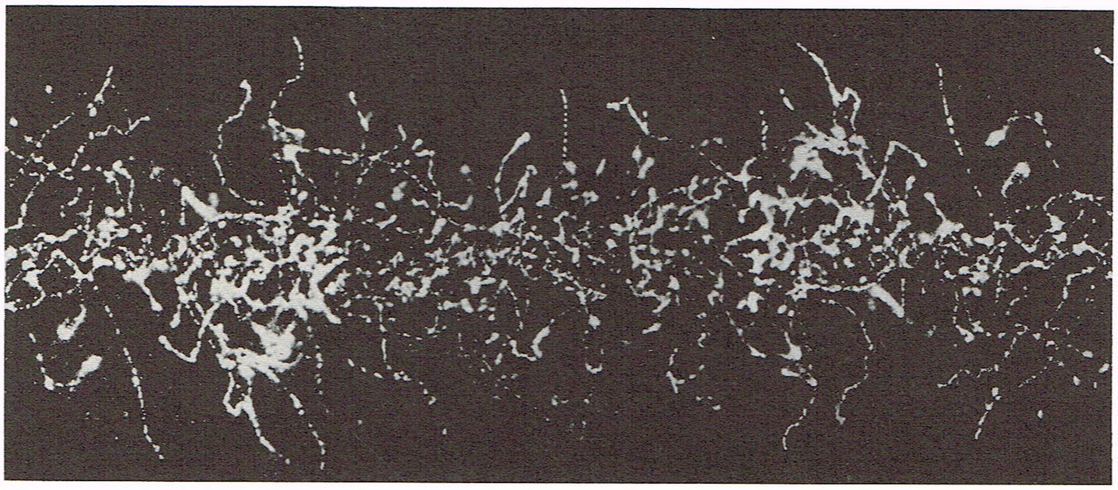
Chambre de Wilson *(coupe verticale)*.

C est un cylindre fermé par un couvercle de verre A, d'une part, et par une membrane mobile M, d'autre part. On met un gaz humide G dans le cylindre. La détente du gaz, due au déplacement brusque du piston P, provoque une condensation de fines gouttelettes sur les ions formés le long de la trajectoire d'une particule rapide. On peut alors photographier (Ph) la trace grâce à un flash puissant L.



Traces de rayons en chambre de Wilson;

les rayons perdent de l'énergie en ionisant les molécules d'air et finissent, à bout de course, par s'arrêter.



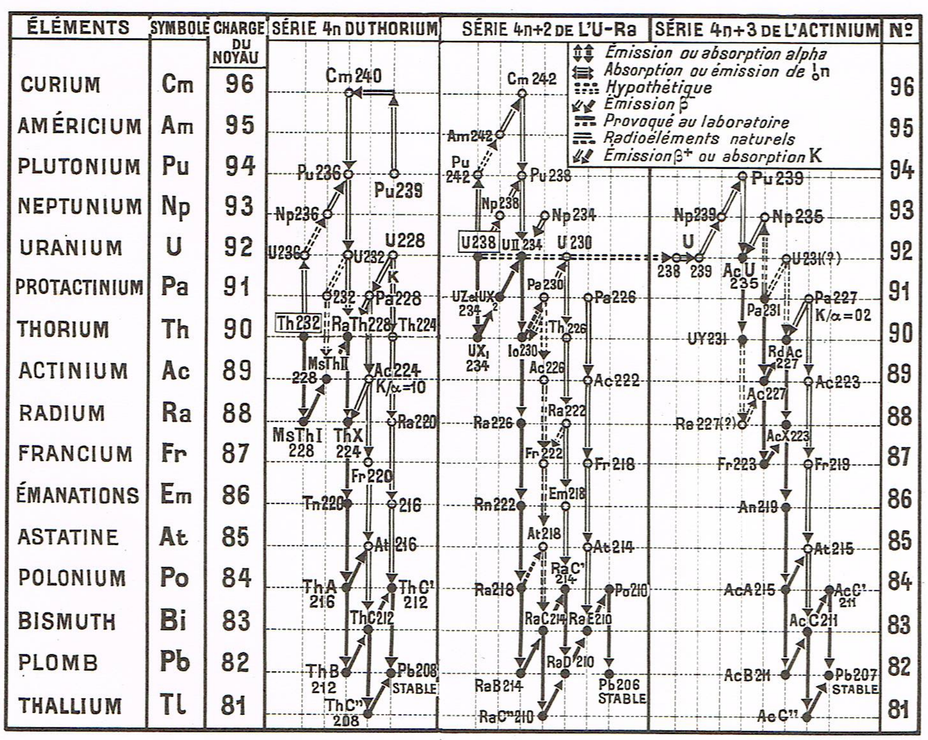
Chambre de Wilson : traces d'électrons libérés par effet photo-électrique le long de la trajectoire d'un mince faisceau de rayons X (invisibles parce que neutres électriquement). Les trajectoires zigzagantes reproduisent le chemin bousculé qu'ont dû se frayer les électrons peu rapides dans la foule des molécules environnantes.

sur les ions, qui jouent le rôle de centres de condensation, avant de remplir la chambre d'un épais brouillard, digne du *fog* londonien. Ainsi les gouttelettes d'eau reproduisent fidèlement les trajectoires des particules qui ont laissé derrière elles des traînées de ions. Un flash permet de prendre la photographie de ces traces, qui ont de quoi vous remplir d'émerveillement; plus prosaïquement, plusieurs photos prises en stéréoscopie permettent de reconstruire et de calculer les trajectoires exactes qu'ont suivies les particules.

Ainsi l'étude de la radio-activité naturelle devait donner le départ, dans les vingt années qui suivirent sa découverte, à toute une série de techniques de plus en plus perfectionnées, permettant même de compter et de voir les particules une à une. **L'ère de la physique corpusculaire était en train de naître.**

le destin des corps radio-actifs

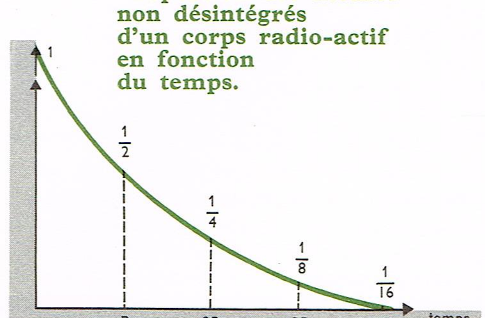
Pendant que s'élaboraient et que se mettaient au point ces techniques, l'étude des éléments radio-actifs commencée par Pierre et Marie Curie se continuait et amenait la reconnaissance d'une quarantaine d'éléments : d'éléments nouveaux comme le radium, ou, le plus souvent, d'isotopes d'éléments stables déjà connus. Peu à peu, on vit plus clair dans cet imbroglio; il y avait des liens de parenté entre les différents corps radio-actifs, et on arriva finalement à les grouper tous en trois familles. Ainsi le radium était un descendant de l'uranium, qui se transformait peu à peu en plomb, élément stable, en passant par l'intermédiaire de treize éléments radio-actifs successifs! La complexité du rayonnement radio-actif n'était ainsi qu'apparente : elle était due à la superposition de plusieurs processus



Les trois familles radio-actives : thorium, uranium-radium, actinium.

Proportion de survivants.

Proportion des atomes non désintégrés d'un corps radio-actif en fonction du temps.



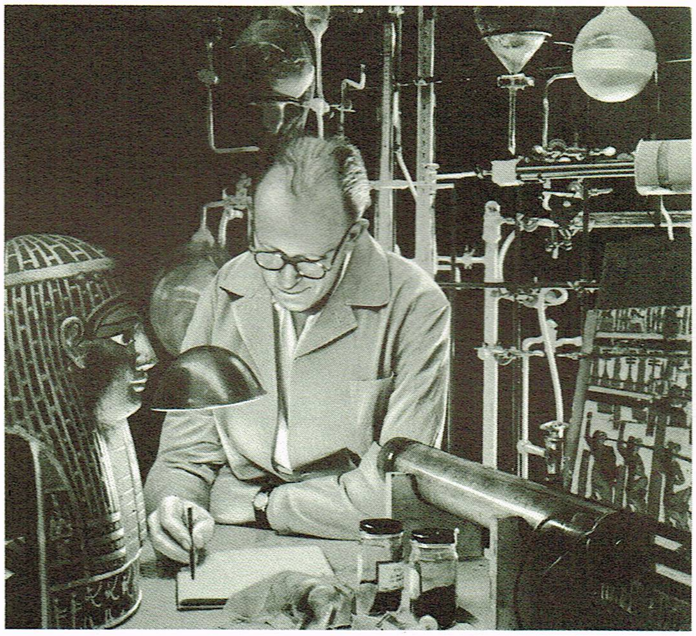
simples; la désintégration de l'atome de radium donnait par exemple naissance à un atome de radon, gaz radio-actif, par l'émission d'une seule particule, transcrite dans la formule simple $Ra_{88}^{226} \rightarrow \alpha_2^4 + Rn_{86}^{222}$, dans laquelle on peut voir que se conservent la somme de charges positives des noyaux (nombre atomique écrit en bas du nom de chaque élément) et celle des poids atomiques (nombre du haut).

L'évolution dans le temps des atomes radio-actifs fit l'objet de nombreuses mesures, rendues possibles par les techniques de comptage. Si le radium, par exemple, émet continuellement des rayons sans avoir l'air de se fatiguer, pourtant il se transforme et disparaît progressivement en tant que radium. Il en est des atomes de radium comme des hommes; mais, alors que chez l'homme le taux de la mortalité est différent à vingt et à soixante-dix ans, pour les atomes de radium il a toujours été trouvé le même, quel qu'en soit l'âge. Foncièrement instables, **les atomes ne vieillissent pas;** en d'autres termes, si, au bout d'un certain temps, la population des atomes de radium, ou de n'importe quel corps radio-actif, a diminué de moitié, au bout d'un autre temps égal, la population des survivants, en moyenne, diminuera également de moitié.

On peut ainsi caractériser entièrement la vie d'un corps radio-actif par ce que l'on appelle sa « période », c'est-à-dire par le temps au bout duquel la moitié de ses atomes se sont désintégrés. Les périodes varient d'ailleurs beaucoup d'un élément radio-actif à un autre; elles s'étendent de 10^{-6} seconde à des milliards d'années. L'uranium de la pechblende a une période de 5 milliards d'années, ce qui explique qu'on puisse encore le trouver dans des roches vieilles comme le monde; le radium a une période plus courte (elle est quand même de 1 600 ans) et se trouve ainsi en faibles proportions par rapport à l'uranium.

Dès 1913, Rutherford suggéra que la propriété qu'ont les corps radio-actifs d'avoir une période invariable, non soumise aux fluctuations du milieu et des conditions climatiques, pourrait constituer une méthode pour établir l'âge des roches contenant des corps radio-actifs. Il suffisait d'établir par exemple la proportion d'uranium, substance mère, par rapport au plomb, substance finale de désintégration. Cette méthode est de loin la plus précise que l'on connaisse de nos jours pour dater l'histoire de la Terre.

Une méthode parallèle, dite « du carbone 14 », a été mise au point plus tard par le physicien américain Libby pour la datation de minéraux organiques ou de plantes remontant à des temps reculés. Le principe en est le suivant. Le gaz carbonique de l'atmosphère contient, à côté du carbone normal C_6^{12}, une faible proportion d'un isotope radio-actif C_6^{14}, dû aux rayons cosmiques qui arrivent très régulièrement dans la haute atmosphère. Ce gaz carbonique a été absorbé durant leur vie par les arbres qui ont formé le charbon, puis le C_6^{14} radio-actif a commencé à se désintégrer et à disparaître lentement. Si l'on suppose que la teneur de l'air en carbone 14 n'a pas varié au cours des temps, on connaît la teneur initiale en carbone 14 de la matière examinée, et on peut, en mesurant la teneur actuelle, en déduire son âge, que ce soit l'âge du charbon des grottes de Lascaux (15 000 ans) ou celui des parchemins de la mer Morte (2 000 ans).



Datation d'œuvres d'art égyptiennes par la méthode du carbone 14.

Pierre et Marie Curie étaient sans doute loin de se douter en 1900 que la radio-activité aurait de telles applications. Ils étaient loin de se douter aussi qu'un jour l'uranium prendrait plus d'importance que leur radium.

de quoi est donc faite la matière nucléaire ?

Inventé par les philosophes grecs de l'Antiquité, l'atome (littéralement l' « insécable ») représentait pour eux la plus petite partie de la matière que l'on pourrait isoler si on la découpait en grains de plus en plus fins; mais cette dernière partie était elle-même indivisible, car comment imaginer que l'on pût diviser la matière à l'infini?

Les chimistes du XIXe siècle crurent-ils avoir trouvé l'atome né de l'imagination des Anciens? Toujours est-il que, moyennant des traitements appropriés, l' « atome » voulut bien se laisser dessaisir d'électrons qu'il contenait, malgré l'apparence d'indivisibilité qu'il se donnait. Et voici que tout un lot de particules, α, β, γ, toutes de grande énergie, pouvait s'échapper de cet édifice. Ces fuites étaient peut-être un accident, mais quelle tentation de forcer la nature à en dire davantage! L'idée de « casser la baraque » ne devait pas cesser d'inspirer la physique corpusculaire de 1910 à nos jours.

le bombardement du noyau atomique par les particules

Rutherford eut le premier l'idée d'envoyer des particules α sur des atomes. Lourdes, elles devaient en effet constituer d'excellents projectiles. Le premier bombardement qu'il entreprit lui révéla, dès 1913, le secret du noyau atomique; en passant au travers d'une feuille d'or, les particules α se heurtaient aux atomes rencontrés, et la façon dont elles ricochaient lui indiqua clairement que l'atome devait se composer de deux parties : au centre, un noyau extrêmement petit — 100 000 fois plus petit que l'atome lui-même — gardien des charges électriques positives et de la presque totalité de la masse de l'atome; autour de ce noyau, quelques électrons qui logent leur petitesse dans un immense espace vide. La conduite des particules était ainsi

Phot. U. S. I. S.

facile à retracer : lourdes, elles n'étaient pas du tout gênées par les poussières d'électrons, réussissaient à atteindre les abords du noyau d'or, mais là, en face d'autres charges positives, elles se faisaient vivement rabrouer.

Le noyau d'un atome radio-actif devait bien être la source des rayonnements α, β, γ, le siège de cette débauche d'énergie que nous ne connaissions pas aux électrons : avec leurs modestes effets chimiques, électriques et lumineux, ils faisaient en effet figure de parents pauvres.

Mais le noyau, défendu par ses charges positives, semblait érigé en une citadelle que les particules α ne pouvaient prendre. Que faire pour en forcer la défense? En choisissant pour cible non des noyaux d'or, mais des noyaux plus légers, on pouvait espérer un résultat : contenant moins de charges positives, ces derniers résisteraient moins bien aux attaques et se laisseraient peut-être casser; quel secret ne livreraient-ils pas alors? Animés d'espoir, les physiciens s'adressèrent alors à tous les rayonnements α connus, passèrent en revue les différents éléments possibles; malgré les difficultés et les limites de leur recherche, ils devaient connaître la joie de la réussite et voir s'ouvrir les frontières du noyau. Les étapes principales et importantes devaient être les suivantes :

1° En 1919, première réaction de chimie nucléaire par Rutherford.

Pour la première fois au monde, le vieux rêve des alchimistes était réalisé; pour la première fois, un élément pouvait être changé en un autre élément. Rutherford réussit, avec des particules α, à changer de l'azote en oxygène (avec un très faible rendement, il est vrai), suivant une réaction nucléaire conservant les charges positives et les masses totales qui s'écrivait $H_2^4 + N_7^{14} \rightarrow O_8^{17} + H_1^1$;

2° En 1932, découverte du neutron par Chadwick.

Chadwick arriva à montrer que, dans certaines réactions nucléaires, il sortait non pas un proton comme dans la première réaction de Rutherford, mais un neutron, c'est-à-dire une particule de même masse que le proton, mais électriquement neutre;

3° En 1934, découverte de la radio-activité artificielle par I. et F. Joliot-Curie.

En bombardant l'aluminium avec des particules α, ils réussirent à créer un isotope radio-actif du phosphore, qui se désintégrait ensuite très rapidement. La désintégration donnait d'ailleurs un e^+, selon un mode inconnu en radio-activité naturelle (l'électron positif e^+, découvert en 1932, est une particule identique à l'électron négatif ordinaire; seule la charge électrique diffère).

Ainsi l'espoir était né de créer des corps radio-actifs de tout genre : il suffisait de se munir d'une artillerie assez variée et souple afin que les projectiles puissent provoquer dans les noyaux des réactions différentes.
Il nous faut insister ici sur la découverte du neutron et les caractéristiques de cette nouvelle particule : la structure du noyau atomique nous paraîtra plus claire.

le neutron

Si l'on regarde la classification périodique des éléments chimiques, on n'est pas sans remarquer — et ceci avait évidemment frappé les physiciens d'avant 1930 — que leurs masses atomiques sont à peu près des multiples entiers de la masse atomique de l'hydrogène. Cette propriété est sans doute à mettre au compte du noyau atomique, puisque la masse des électrons planétaires, négligeable devant celle du proton, ne peut guère (1) avoir une influence dans ce cas; il est tentant de supposer que le noyau d'hydrogène, le premier de la série (d'où son nom de « proton » du grec *protos*, « premier »), joue un rôle particulier dans cette suite.

Cette hypothèse se fit d'autant plus plausible qu'on s'aperçut que des éléments comme le chlore, aux masses atomiques récalcitrantes (35,5), révélaient au spectrographe de masse des noyaux de masses différentes et entières (pour le chlore, 35 et 37). Comment alors imaginer la formation d'un noyau, comme par exemple celui d'hélium He_2^4, à partir d'éléments plus simples? On peut en particulier retenir — comme le montre



1re possibilité :
He_2^4 = 4 protons + 2 électrons;



2e possibilité :
He_2^4 = 2 protons + 2 neutrons.

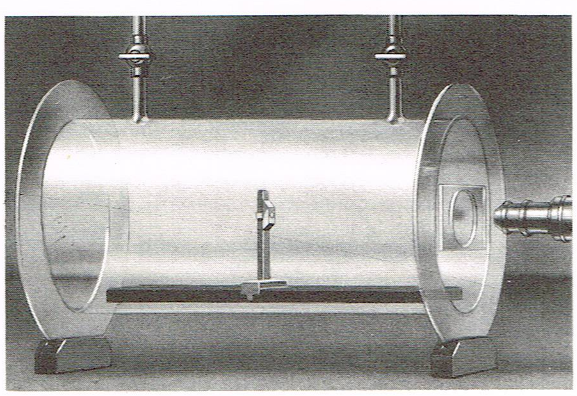
le schéma — deux possibilités qui assurent le nombre de charges positives (deux) et la masse atomique (quatre) de ce noyau : la première de ces éventualités semblait même la plus plausible, puisque aussi bien la radio-activité naturelle nous révélait des électrons qui sortaient du noyau atomique, jusqu'au jour où Chadwick mit en évidence l'existence du neutron.

L'histoire de cette découverte vaut la peine d'être racontée, car elle illustre assez bien comment, à partir d'un fait expérimental bizarre, on arriva à suspecter et ensuite à prouver l'existence du neutron, qui n'était jusqu'alors qu'une hypothèse fantaisiste.

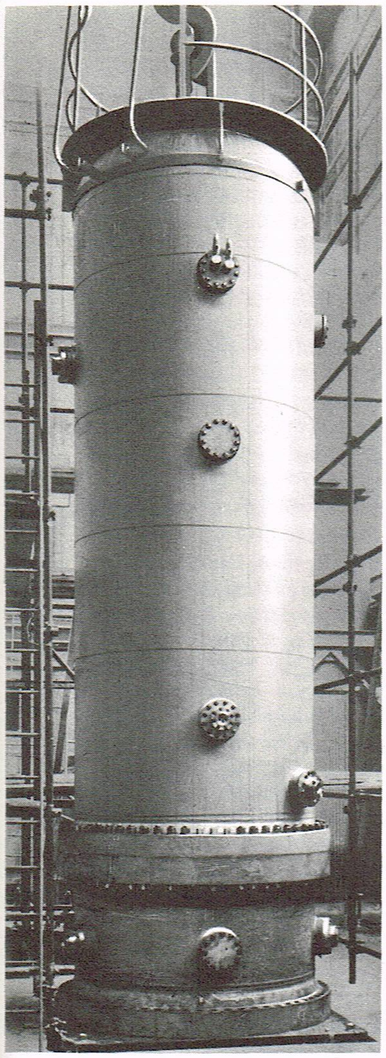
Or donc, Bothe et Becker, en 1930, en envoyant des particules sur du béryllium, décelèrent avec des compteurs de Geiger l'émission d'un rayonnement neutre. Ils l'interprétèrent évidemment comme un rayonnement γ, qu'ils voulaient justement mettre en évidence dans des réactions nucléaires. I. et F. Joliot-Curie, reprenant l'expérience, montrèrent que ce rayonnement était très énergique, capable d'éjecter des protons d'une substance hydrogénée. Pour que des rayons γ pussent ainsi bousculer des protons, il fallait leur supposer une énergie d'environ 50 MeV, bien supérieure aux quelques méga-électron-volts jusque-là observés dans la radio-activité naturelle.

Frappé de cette bizarrerie, Chadwick, collaborateur de Ruther-

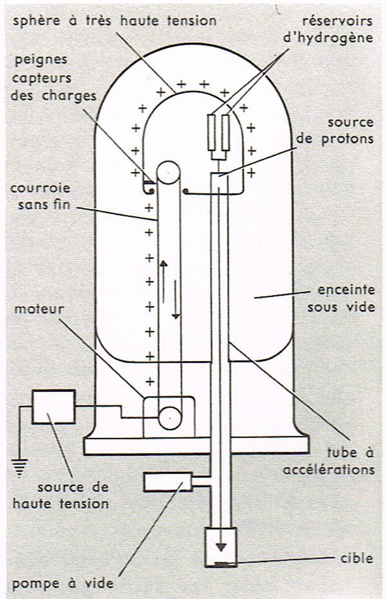
L'appareil qui permit à Rutherford de réaliser le vieux rêve des alchimistes. Le petit bras mobile à l'intérieur du cylindre porte le radium émetteur de particules qui transmuteront l'azote en oxygène. Le phénomène était observé sur un écran fluorescent par le microscope visible à droite.



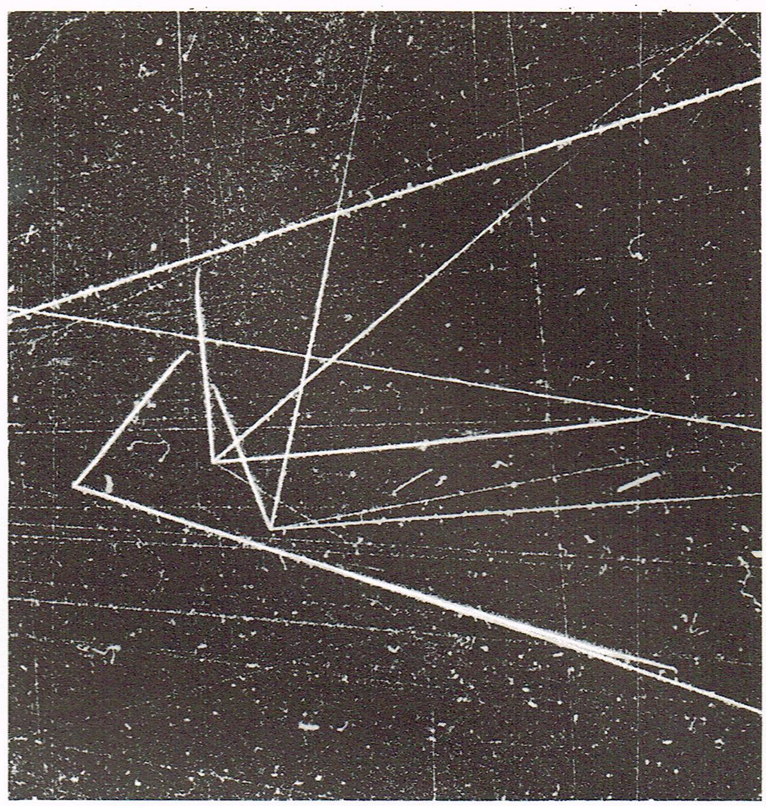
(1) On se rappelle que l'électron est 1 840 fois plus léger que le proton, noyau d'hydrogène.



L'accélérateur Van de Graaff du Centre atomique de Saclay.



Croquis schématique d'un appareil van de Graaff.



Ci-dessus, **désintégration de noyaux de gaz produite dans la chambre de Wilson par des protons et des neutrons du synchrocyclotron de Berkeley.**

ford, partit sur une autre piste, suspectant une particule nouvelle. Il monta une expérience dans laquelle il put déterminer la répartition des énergies dans la réaction nucléaire et mesurer ainsi la masse de la particule (1) neutre qui occasionnait le mystérieux rayonnement; il la trouva légèrement supérieure à celle du proton. Le neutron, symbolisé par n_0^1, était découvert, et la réaction nucléaire qui l'avait produit s'écrivait simplement $H_2^4 + Be_4^9 = C_6^{12} + n_0^1$.

Il fallait bien conclure que le **noyau était formé d'un simple assemblage de protons et de neutrons,** qu'on appelle encore indistinctement « nucléons ». Leur nombre se déduisait automatiquement du numéro atomique (nombre de protons) et de la masse atomique (nombre de nucléons) de l'élément; ainsi

hydrogène	H_1^1	: 1 proton;
hélium	H_2^4	: 2 protons + 2 neutrons;
oxygène	O_8^{16}	: 8 protons + 8 neutrons;
uranium 238	U_{92}^{238}	: 92 protons + 146 neutrons.

D'ailleurs, l'hypothèse que le noyau pouvait contenir des électrons était en contradiction avec la nature ondulatoire de la matière, précisée par L. de Broglie. L'électron de la radioactivité β venait bien du noyau, mais il n'y préexistait pas; il y avait véritable création; d'ailleurs, on devait montrer plus tard (1951) que le neutron isolé n'était pas stable, mais se désintégrait en donnant un proton et un électron. S'il pouvait y avoir des transformations entre noyaux atomiques, **il y avait également transformation de particules entre elles.**

(1) La mesure se faisait selon la technique du bilan nucléaire, conséquence du principe d'équivalence masse-énergie.

le perfectionnement de l'artillerie nucléaire : le « Van de Graaff » et le cyclotron

Si les particules α de la radio-activité naturelle pouvaient casser les noyaux légers — et l'on vient de voir les merveilleux résultats qui sont dus à cette découverte —, elles n'étaient d'aucune utilité pour des noyaux plus lourds : pas de formation de nouveaux éléments, pas de création d'isotopes radio-actifs inconnus; ces noyaux restaient intacts. Pour exploiter cet autre domaine, il fallait d'autres moyens. Certes, après la découverte du neutron on disposa d'une particule qui devait être prometteuse, puisque, neutre, elle pouvait s'approcher du noyau plus facilement qu'une particule α flanquée de ses deux charges positives. On réussit, en effet, à en faire un agent de réactions nucléaires, et le neutron devait atteindre plus tard la notoriété avec la bombe atomique.

Cependant, dès 1930, les physiciens s'étaient aussi jetés dans une autre direction. Si les particules α de la radio-activité naturelle avaient de telles limites, pourquoi ne pas les fabriquer (on savait enlever deux électrons à un atome d'hélium) et les accélérer artificiellement? Oh certes! le problème n'était pas d'une simplicité enfantine : il faut se rappeler que la nature donne déjà des particules α, dont la vitesse correspond à celle que donnerait une différence de potentiel de quelques millions de volts! Et pourtant, quel moyen merveilleux n'aurait-on pas? On pourrait accélérer des noyaux d'hélium, mais aussi des

protons et d'autres particules, et non se contenter de ce que veut bien donner la nature! On pourrait obtenir un nombre de particules supérieur aux maigres apports de la radio-activité naturelle!

Ainsi germa l'idée de construire des accélérateurs et c'est à leur réalisation que l'on doit l'étude systématique de toute réaction nucléaire imaginable, de toute création d'isotopes inconnus dans la nature, stables ou instables. Ces outils de recherche ont constitué d'ailleurs une branche d'activité qui s'est continuée jusqu'à nos jours, se perfectionnant et se développant, les chercheurs essayant d'atteindre des énergies toujours supérieures.

Si l'on voulait avoir une particule possédant une énergie de 1 MeV, il fallait par exemple l'accélérer par une différence de potentiel électrique d'un million de volts. On soupçonne les difficultés d'une telle entreprise, les problèmes d'isolement électrique n'étant pas les moindres à surmonter. Plusieurs types d'appareils furent construits, dont l'un des plus connus est le « Van de Graaff », du nom du physicien belge qui l'inventa. Plus un conducteur électrique est à potentiel élevé, plus il contient de charges électriques; pour atteindre des tensions aussi fortes que le million de volts, l'idée de Van de Graaff fut d'amener les charges une à une sur un conducteur isolé, et ce au moyen d'un tapis roulant isolant. De même que des tapis roulants transportent des valises, le tapis roulant de Van de Graaff transporte des charges électriques; il les prend à un endroit où un ensemble transformateur-redresseur de 10 000 à 30 000 V se charge de les lui fournir et les amène à l'intérieur d'une sphère métallique, où un peigne fin conducteur les cueille au passage et les met au contact de la sphère. Se repoussant les unes les autres, les charges électriques s'enfuient le plus loin possible et vont élire domicile sur la surface externe de la sphère. Ainsi, peu à peu, les charges électriques, positives par exemple, s'accumulent sur la sphère, dont le potentiel augmente de plus en plus. Si une source de protons (ou d'autres ions) est placée à l'intérieur de la sphère, elle émet des particules qui, violemment repoussées par la sphère, seront éjectées avec la vitesse et l'énergie déterminées par la différence de potentiel obtenue. Guidées à la sortie, elles seront alors envoyées sur toute cible au choix, où elles pourront donner lieu à des réactions nucléaires.

En fait, même en utilisant des surfaces sphériques bien polies, pour éviter les décharges électriques, et en faisant le vide le plus poussé dans l'appareil, l'isolement électrique a des limites, et il est difficile avec un « Van de Graaff » d'obtenir des protons pourvus d'une énergie supérieure à 10 MeV. C'est pour tourner ces difficultés qu'il pressentait, que Lawrence conçut en 1931 le cyclotron, appareil bien différent: au lieu d'appliquer en une seule fois, à la particule, la différence de potentiel requise, il pensa à la lui administrer par petits coups successifs. Le cyclotron est essentiellement une boîte cylindrique métallique, espèce de boîte de camembert qui aurait quelque chose comme 50 cm de diamètre. Cette boîte est coupée en deux parties égales et les surfaces de séparation forment des électrodes auxquelles une tension électrique est appliquée. Les particules peuvent circuler dans la boîte (où l'on a fait le vide), et un champ magnétique perpendiculaire à celle-ci les oblige à tourner en rond. Quand elles traversent la surface de séparation des deux demi-boîtes, elles sont soumises à la tension électrique appliquée aux électrodes et sont accélérées. Quand elles ont parcouru une demi-circonférence et qu'elles changent à nouveau de demi-boîte, elles peuvent encore être accélérées si entre-

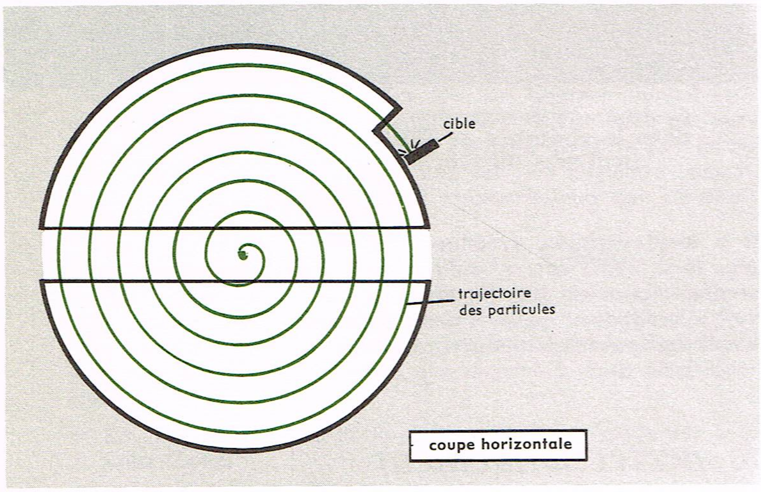
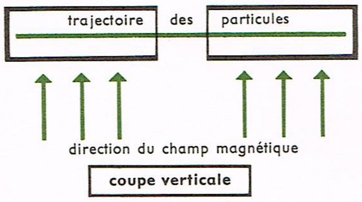
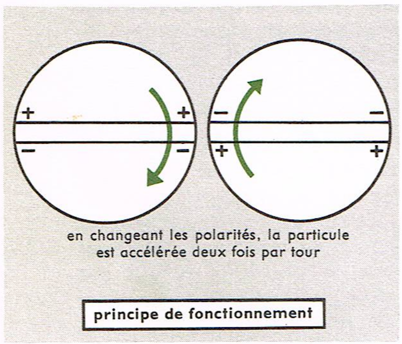
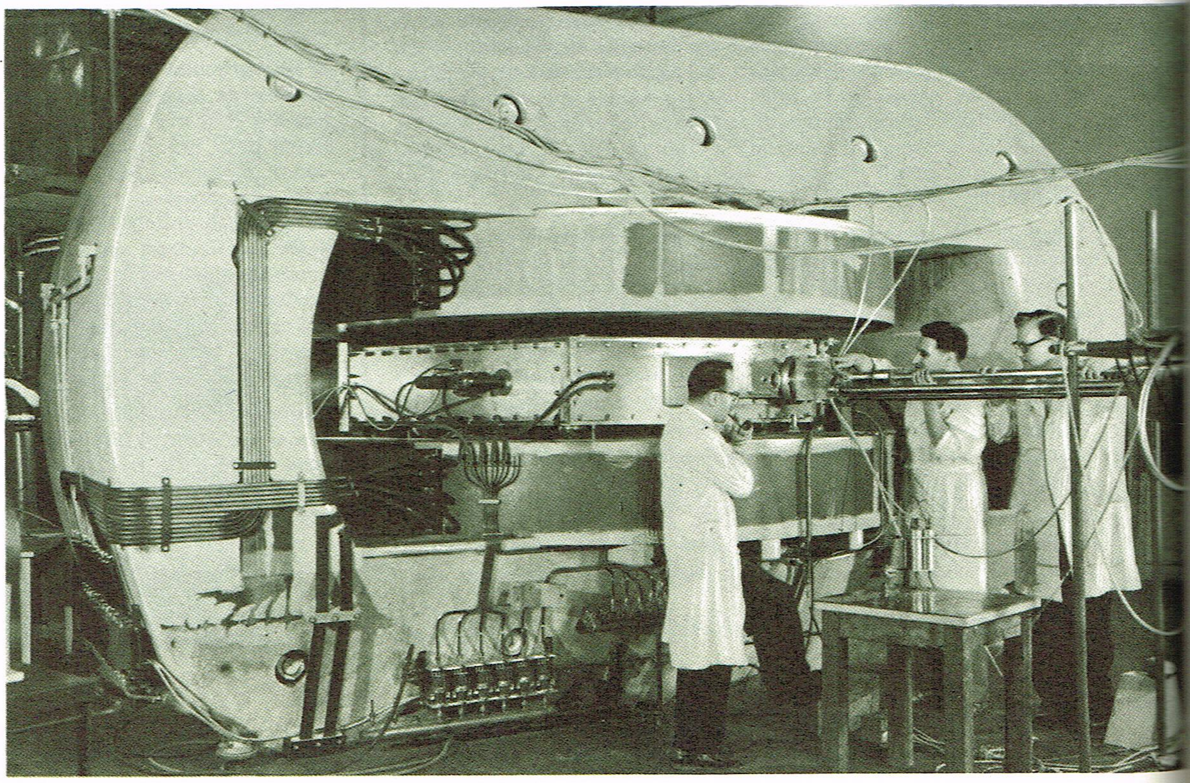


Schéma du cyclotron.







Le cyclotron du Centre d'études atomiques de Saclay.

temps on s'est arrangé pour changer la polarité des électrodes. Toute l'astuce de l'appareil réside dans le fait suivant : si le champ magnétique est gardé constant (ce qui est facile à réaliser) à chaque nouvelle accélération les particules décrivent un cercle plus grand, mais en fait, allant plus vite, elles mettent toujours le même temps pour parcourir une demi-circonférence. Ainsi le changement de polarité des électrodes peut être obtenu par une tension alternative de fréquence convenable. Certes, les particules vont vite et font plus d'un million de tours par seconde et il faut des tensions de haute fréquence, mais on sait produire de telles tensions. L'essentiel est que la fréquence reste constante pendant toute l'accélération. Ainsi accélérées à chaque tour et emportées par leur vitesse grandissante, les particules décrivent une spirale et, en fin de course, se précipitent toutes seules sur une cible préparée à leur intention.

Le cyclotron peut accélérer des particules jusqu'à leur donner une énergie de 40 MeV environ, les énergies supérieures lui étant inaccessibles parce qu'alors le temps de révolution des particules ne reste plus constant (nous en verrons les raisons plus loin). Le cyclotron devra alors être modifié pour donner naissance au synchro-cyclotron; il aura pourtant suffi à casser les noyaux les plus lourds, et la collection d'isotopes, stables et instables, en sera complétée.

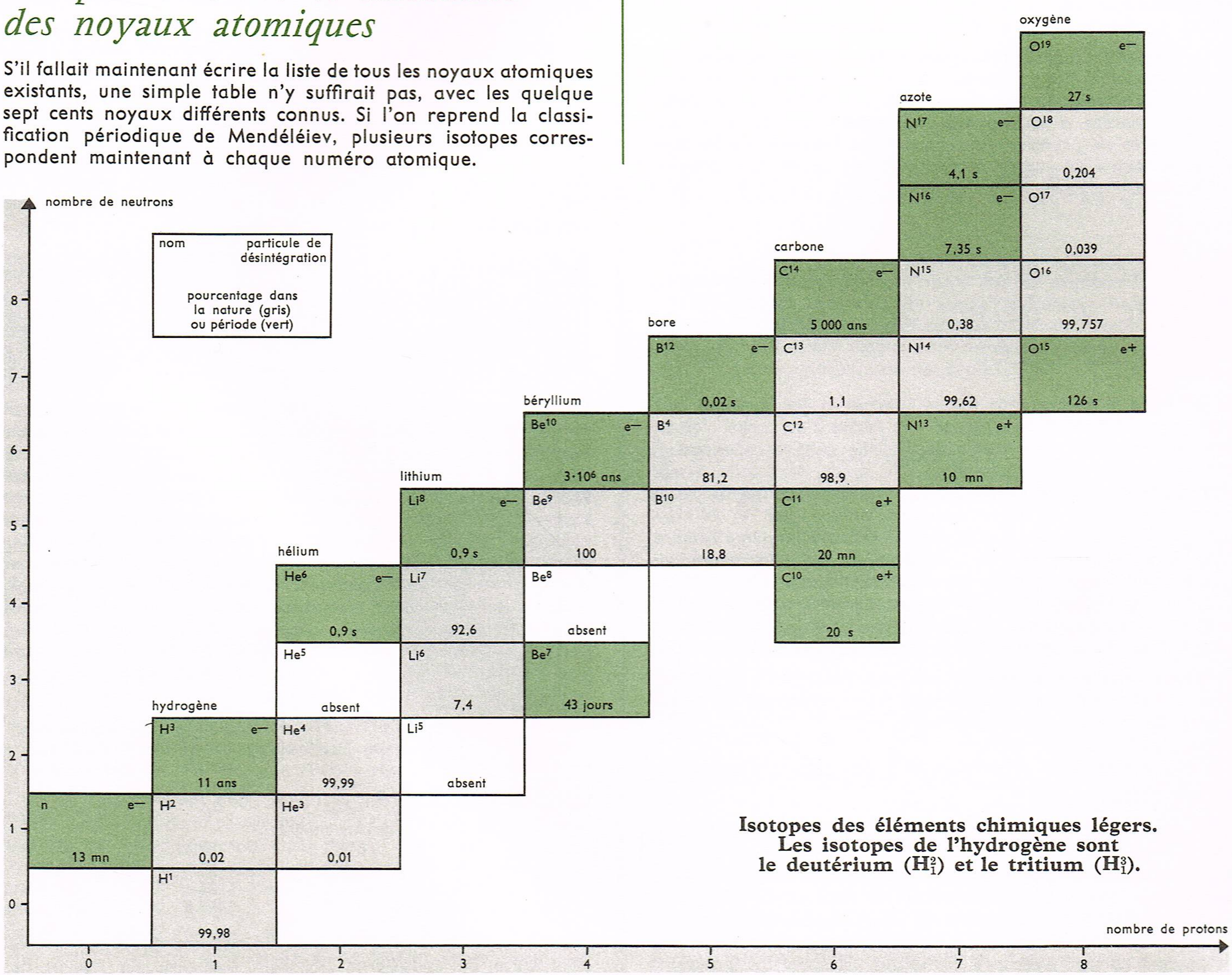
isotopes. Stabilité et instabilité des noyaux atomiques

S'il fallait maintenant écrire la liste de tous les noyaux atomiques existants, une simple table n'y suffirait pas, avec les quelque sept cents noyaux différents connus. Si l'on reprend la classification périodique de Mendéléiev, plusieurs isotopes correspondent maintenant à chaque numéro atomique.

Ces isotopes forment une classe d'atomes frères, car ils possèdent le même nombre de protons et, partant, le même nombre d'électrons; seul celui de neutrons diffère. Même arrangement des électrons, donc mêmes propriétés chimiques, et ce n'est pas par hasard que ces isotopes ne furent pas découverts par les chimistes, dont les procédés ne peuvent permettre leur mise en évidence : il faut faire appel à d'autres méthodes, fondées sur la légère différence de masse qu'ils possèdent entre eux. Si leur identification peut se faire assez facilement sur de petits échantillons par spectrographie de masse (méthode qui permit à J. J. Thomson de les découvrir), leur séparation à un échelon important pose souvent de sérieux problèmes. Un cas typique est celui de la séparation des deux isotopes de l'uranium, U^{235} et U^{238}, qui demande de très gros moyens financiers.

Les isotopes sont formés de noyaux de deux sortes. Certains, tel le deutérium (isotope de l'hydrogène, qui entre dans la composition de l'eau lourde), sont stables, pouvant se conserver indéfiniment. D'autres, tel le tritium, sont instables et radioactifs, caractérisés par des périodes qui s'étalent de la fraction de seconde au million d'années. Ils se désintègrent en général en d'autres éléments voisins stables.

Quelles lois président donc à la stabilité des noyaux? Une première constatation se dégage du tableau ci-contre : les noyaux



Isotopes des éléments chimiques légers. Les isotopes de l'hydrogène sont le deutérium (H_1^2) et le tritium (H_1^3).

stables légers contiennent un nombre à peu près égal de protons et de neutrons. Le proton est stable, le deutérium (neuton + proton) l'est aussi. Quant à l'He^2 (deux protons), il n'existe même pas : tout se passe comme si les forces nucléaires n'étaient pas assez puissantes pour maintenir ensemble deux protons et contrebalancer l'effet de la répulsion électrostatique due aux charges électriques, alors qu'elles assurent la cohésion du proton et du neutron dans le deutérium. Au contraire, si entre deux protons un neutron s'interpose, le noyau formé (He^3) est stable. Ainsi, dans les noyaux stables le nombre de neutrons augmente de pair avec celui des protons, pour empêcher ces derniers de se repousser trop violemment et de se séparer; et même, plus le noyau grossit, plus les forces de dislocation doivent augmenter, car la stabilité n'est assurée que si le nombre de neutrons excède de plus en plus celui de protons. Cet état de choses se comprend assez bien si l'on établit une analogie entre un noyau et une goutte d'eau, que les physiciens ont, d'ailleurs avec quelque succès, prise comme modèle de la structure nucléaire.

Il est plus facile d'avoir des gouttes fines que grosses, et chacun sait qu'une goutte de liquide ne peut dépasser une certaine dimension sans se briser. Ainsi en va-t-il des noyaux, que le nombre de neutrons de plus en plus nombreux n'arrive plus à cimenter avec succès. C'est à partir du numéro atomique Z = 86 (radon) que cette stabilité disparaît. Commence alors le groupe des corps radio-actifs naturels, qui s'étend jusqu'à l'uranium (Z = 92). Au-delà, les transuraniens, inconnus dans la nature, ont été créés artificiellement jusqu'à Z = 102.

Si ces remarques assez qualitatives ont permis de faire des hypothèses sur la nature des forces nucléaires, une plus grande partie des renseignements quantitatifs est venue de l'application du principe d'équivalence masse-énergie d'Einstein, qui joue un rôle de premier plan dans toute l'histoire de la physique nucléaire en même temps qu'il en reçoit de multiples vérifications.

équivalence masse-énergie. Énergie de liaison des noyaux : défaut de masse

Le principe d'équivalence masse-énergie a été formulé par Einstein en 1905. Il consiste en ceci : la masse d'un corps, c'est-à-dire cette quantité responsable de l'inertie dans le mouvement, et responsable, avec la pesanteur, du poids, bref cette masse définie par la mécanique équivaut à une énergie interne de ce corps. La relation d'équivalence est définie par la célèbre formule $\mathbf{E = mc^2}$, où c désigne la vitesse de la lumière (300 000 km/s). Que l'on se méfie tout de suite des complications en même temps que des fausses interprétations : la masse m d'un corps et l'énergie $E = mc^2$ correspondante sont à entendre pour un corps au repos; s'il est en mouvement, il se surajoute à cette énergie de masse une énergie cinétique propre.

On peut comparer par exemple, pour une masse de 1 kg animée d'une vitesse déjà bien grande de 1 km/s, l'énergie de masse à l'énergie cinétique correspondante :

énergie de masse :

$$E = mc^2 = 1\,000 \text{ g} \times 3.10^{10} \text{ cm/s} \times 3.10^{10} \text{ cm/s} = 9.10^{23} \text{ ergs;}$$

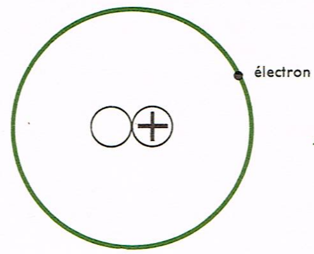
énergie cinétique :

$$T = \frac{1}{2}mv^2 = \frac{1}{2} \times 1\,000 \text{ g} \times 10^5 \text{ cm/s} \times 10^5 \text{ cm/s} = 5.10^{12} \text{ ergs.}$$

Ainsi il n'existe aucune commune mesure avec nos habituelles conceptions : la masse est un réservoir immense d'énergie et les vues d'Einstein se sont trouvées amplement confirmées par toute la physique nucléaire, où les interactions et les mutations de noyaux se font avec des variations sensibles de masses et donc avec une libération importante d'énergie cinétique ou de chaleur.

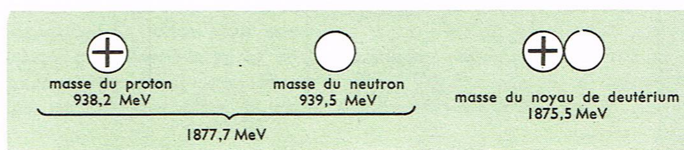
Sous cet angle, il vaut la peine de s'intéresser à la formation des noyaux à partir de leurs éléments constitutifs et d'étudier la répartition des masses. Selon les vues d'Einstein, il est préférable de mesurer directement les masses en unités d'énergie, et plus exactement en méga-électron-volts, unité qui nous est maintenant familière, quitte à donner tout de suite les correspondances, dont il est bon de temps en temps de se rappeler l'ordre de grandeur : 1 g = $5{,}61 \times 10^{26}$ MeV; masse d'un proton = $1{,}672 \times 10^{-24}$ g = 938,2 MeV.

Il est évident que la stabilité du noyau ne peut venir que de forces qui maintiennent les nucléons collés ensemble, car comment pourrait-on expliquer autrement que le proton et le neutron de l'atome de deutérium par exemple restent confinés dans ce tout petit volume du noyau, alors qu'un espace



Atome de deutérium.

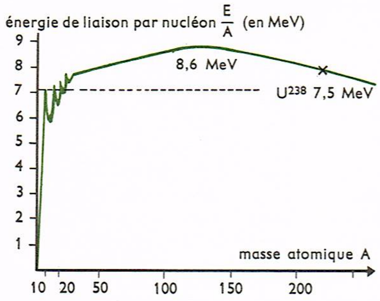
100 000 fois plus grand leur est généreusement alloué ? Solidement attachés l'un à l'autre, ils ne se laisseront séparer et rendre à la liberté que fortement violentés, à coups de bombardement nucléaire. L'énergie qu'on aura alors dépensée devra, bien entendu, se retrouver automatiquement dans la différence d'énergie des deux systèmes proton + neutron libres et noyau de deutérium. Si les idées d'Einstein sont justes, elle doit se retrouver dans une différence de masse, et c'est bien ce que l'on constate, car le bilan s'établit ainsi :



La différence de masse de 2,2 MeV mesure l'énergie de liaison de ces deux nucléons dans le noyau de deutérium, énergie minimale qu'il faut fournir si on veut les séparer. De la même façon on peut obtenir l'énergie de liaison d'un noyau quelconque, c'est-à-dire l'énergie qu'il faudrait pour le disloquer et en éparpiller tous les nucléons. Tous les noyaux, en effet, présentent ce défaut de masse, comme l'avaient déjà noté les chimistes : mesurée par rapport à l'hydrogène, la masse atomique d'un élément contenant A nucléons est à peu près A, mais en fait légèrement inférieure; nous en comprenons maintenant la raison. Quand on pense que ce défaut de masse seul permet la libération immense d'énergie nucléaire que chacun sait, il est hallucinant de constater qu'il ne constitue que la centième partie de la masse totale!

élément	nombre de nucléons	masse atomique	énergie totale de liaison (MeV)	énergie de liaison par nucléon (MeV)
H^1	1	1,000 (par convention)	—	—
He^4	4	3,972	28,3	7,72
Li^6	6	5,969	32,0	5,33
O^{16}	16	15,870	127,5	7,98
Fe^{56}	56	55,509	487,9	8,72

Pour comparer la stabilité relative des divers noyaux, il est intéressant de considérer l'énergie moyenne de liaison des



Cette courbe donne tout de suite une idée de la stabilité des divers noyaux, les plus stables étant ceux où l'énergie de liaison est la plus élevée.

nucléons, c'est-à-dire l'énergie totale de liaison du noyau divisée par le nombre de nucléons qu'il renferme. Plus elle est grande et plus la stabilité est assurée. En suivant la liste des masses atomiques croissantes, on voit qu'elle est d'abord irrégulière, donnant à la première partie de la courbe ci-contre une allure brisée : les noyaux de masse atomique 4, 8, 12, 16 (He^4, Be^8, C^{12}, 0^{16}) sont bien plus stables que leurs voisins; mais, dans l'ensemble, l'énergie de liaison augmente pourtant lentement, passant de 2 MeV pour le deutérium à plus de 8 MeV pour des noyaux de masse atomique comprise entre 30 et 60. Puis elle redescend lentement. On voit déjà par cette courbe que l'on pourrait en théorie récolter beaucoup d'énergie si l'on accolait des noyaux légers (domaine de la *fusion*, pour $A < 60$) ou si l'on cassait des noyaux lourds (domaine de la radio-activité naturelle et de la *fission*, pour $A > 60$).

De toute façon, l'une des caractéristiques essentielles d'une réaction nucléaire, on le voit bien, sera son bilan énergétique, donné par la différence de masse entre les noyaux au départ et à l'arrivée. Positive, la réaction libérera de l'énergie; négative, il faudra lui en fournir pour qu'elle se fasse. Du second type relève la toute première réaction nucléaire réussie par Rutherford; du premier, celle qui fut obtenue d'abord par Cockcroft et Walton avec les accélérateurs naissants :

$$He_2^4 + N_7^{14} = 0_8^{17} + H_1^1 - 1{,}15 \text{ MeV};$$
$$Li_3^7 + H_1^1 = 2\ He_2^4 + 17{,}7 \text{ MeV}.$$

Pour donner une idée cohérente de la structure nucléaire, on imagina des comparaisons qui puissent rendre compte des diverses propriétés expérimentales. On a déjà mentionné le modèle de la goutte, qui connut quelque succès pour expliquer l'instabilité des noyaux lourds; on indiqua le modèle des couches, reproduisant un peu les couches électroniques de l'atome, pour expliquer la périodicité rencontrée dans les énergies de liaison des noyaux légers. Mais aucune comparaison n'a jamais été pleinement satisfaisante. **Le véritable secret des forces nucléaires n'est pas encore percé, et le mystère reste entier.** Si l'on connaît parfaitement les lois qui régissent les forces électriques, on n'a pas encore la moindre idée de la loi de base qui expliquerait le comportement de ces forces beaucoup plus importantes que sont les forces nucléaires. Depuis la découverte de la radio-activité, personne n'a su trouver la clef de l'énigme. Personne ne sait même aujourd'hui quelle serait la bonne direction à suivre dans les recherches. Peut-être faudra-t-il qu'un jour une idée géniale bouleverse les traditions de la physique?

Cette ignorance n'empêchera pas d'ailleurs l'extension des applications de cette énergie nucléaire.

fission et fusion nucléaires, réservoirs d'énergie

A l'échelle d'un atome, les énergies mises en jeu sont de l'ordre de l'électron-volt pour une réaction chimique et du million d'électron-volts (MeV) pour une réaction nucléaire. A notre échelle, cette énorme différence est en général bien atténuée, et les raisons en sont simples : la réactivité chimique est quasi instantanée, mais la radio-activité s'étale dans le temps, ne devenant parfois appréciable qu'après des millions d'années; le rendement d'une réaction chimique est pratiquement de 100 p. 100, mais celui d'une réaction nucléaire est souvent infime. Dès la découverte en 1919 de la première réaction nucléaire, on entrevit à la fois les possibilités énergétiques et les difficultés de ces jeux de noyaux. On se rendit vite compte aussi que les combinaisons les plus faciles à réaliser avec un bon rendement devraient être celles qui accolent des noyaux légers, puisque la répulsion électrostatique est alors plus faible. De plus, certaines de ces réactions libéreraient une grande quantité d'énergie, telles celles qui permettraient de transformer de l'hydrogène en hélium.

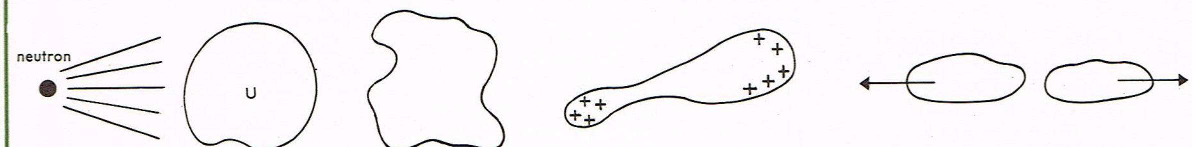
Comme l'hydrogène est une matière première abondante dans notre monde, ne fût-ce que dans l'eau, la « fusion » de noyaux légers avec un rendement appréciable devrait ouvrir des perspectives énergétiques immenses; posé depuis longtemps, le problème n'est, en fait, pas encore résolu.

C'est au contraire un phénomène de découverte plus récente (1939) qui permit pour la première fois l'utilisation et la domestication de l'énergie nucléaire. Dans certaines conditions, l'uranium ainsi que d'autres noyaux très lourds peuvent se casser en deux — on dit alors qu'ils subissent la fission — et donner naissance à deux noyaux de masse intermédiaire. Ces noyaux produits sont plus stables et la cassure se fait avec transformation en énergie d'une partie de la masse du noyau d'uranium.

La question capitale qui nous intéresse ici est celle des moyens qui ont permis d'obtenir de façon abondante la fission des noyaux lourds. Après avoir vu comment ce délicat problème de rendement a été résolu pour la fission, nous l'aborderons pour la fusion, dans toute sa complexité.

la fission nucléaire dans le silence des laboratoires

La fission nucléaire est l'un des exemples, assez fréquents en physique, d'une découverte à l'état pur : personne ne la prévoyait, personne même n'y songeait, et c'est en procédant à d'autres recherches qu'elle fut trouvée. Son histoire commence en 1934. La physique nucléaire était alors toute orientée vers l'obtention de réactions engendrées par les particules α de la radio-activité naturelle. L'exploration de ce domaine devait conduire en particulier à la découverte des neutrons. De là naquit l'idée de se servir de ces neutrons pour produire des réactions nucléaires. Un nouveau domaine pouvait s'ouvrir, riche de promesses. En effet, entre des neutrons sans charge électrique et des particules α doublement chargées, les premiers devaient sans conteste s'approcher plus facilement des noyaux et constituer de meilleurs agents nucléaires. Fermi, en 1934, commença en ce sens une série d'expériences, étudiant l'action de ces neutrons sur différents noyaux. Leurs effets sur l'uranium l'intéressèrent au plus haut point; en s'adressant au noyau le plus lourd connu, au dernier de la classification périodique, il avait l'espoir, par addition de neutrons, de produire d'autres noyaux inconnus plus lourds encore et de prolonger la classification périodique par des éléments transuraniens, de nombre



Déformation du noyau et secousses violentes sous l'effet de choc du neutron.

Les charges électriques favorisent la déformation du noyau.

Le noyau se casse en deux morceaux qui emportent l'énergie libérée sous forme d'énergie cinétique.

Schéma de la fission nucléaire : on peut y voir une certaine analogie avec les cellules vivantes qui se divisent pour se reproduire, avec également des gouttes d'eau sur une plaque brûlante, qui se secouent et se sectionnent avant de disparaître.

atomique supérieur à 92. Il pensa y être arrivé, mais les résultats de ses expériences délicates étaient complexes et difficiles à interpréter. Quelques rares chercheurs continuèrent dans cette voie, obtenant d'autres résultats complexes; cela jusqu'en 1938, où les Allemands Hahn et Strassmann, experts en analyse chimique, mirent en évidence la présence d'un isotope du baryum dans les produits de réaction. Le premier pas vers la découverte de la fission nucléaire de l'uranium était franchi : sous l'effet du choc des neutrons, le noyau d'uranium donnait peut-être des transuraniens, mais, chose à laquelle personne n'avait pensé, il pouvait aussi se casser, se divisant en deux noyaux de masse intermédiaire. Ainsi, par exemple :

$$U^{235}_{92} + n^1_0 \longrightarrow Ba^{142}_{56} + Kr^{94}_{36}.$$

Plusieurs réactions prenaient, en fait, naissance, rendant complexe le phénomène, mais la fission nucléaire existait réellement, comme d'autres expériences devaient vite le confirmer.

On comprend l'intérêt suscité par cette découverte quand on se reporte à la courbe des énergies moyennes de liaison des nucléons pour les divers noyaux. C'est une différence de 1 MeV environ qui existe entre l'uranium et ses produits de fission; autrement dit, quand un noyau d'uranium avec ses quelque deux cents nucléons se brise et se divise en deux morceaux de grosseur comparable, c'est une énergie de 200 MeV qui doit se trouver libérée. Les réactions nucléaires connues jusque-là transformaient, par exemple par addition de protons, un noyau en un autre noyau de nombre atomique voisin, et l'énergie obtenue par différence de masse pouvait tout au plus atteindre 20 MeV. La fission du lourd noyau d'uranium doit libérer au contraire une énergie au moins dix fois plus grande.

Le rendement par noyau est donc bien meilleur. On avait une raison d'espérer qu'une libération d'énergie intéressante à notre échelle serait plus facile avec la fission qu'avec toute autre réaction nucléaire. A quoi correspondrait cette énergie de 200 MeV par noyau si l'on réussissait à casser tous les noyaux d'un morceau d'uranium? **Un gramme d'uranium libérerait alors l'énergie fantastique de 19 millions de kilocalories, autant que la combustion de 2,5 t de charbon!**

Ainsi les conséquences pratiques de cette fission nucléaire toute nouvelle venue pouvaient se révéler grandioses, et aussi terribles. C'était en 1939, au début de la Seconde Guerre mondiale, et les recherches devinrent, très vite et partout, secrètes. Avec l'aide efficace des physiciens exilés d'Europe, elles devaient aboutir aux Etats-Unis au fonctionnement, en 1942, à Chicago, de la première pile atomique, réalisée par Fermi, puis à la construction de la première bombe atomique.

Comment des quantités infimes de baryum, curiosité de laboratoire, trouvées par la sagacité de Hahn et Strassmann, a-t-on pu passer à ce déchaînement d'énergie nucléaire que le monde apprit avec stupeur le 6 août 1945 ?

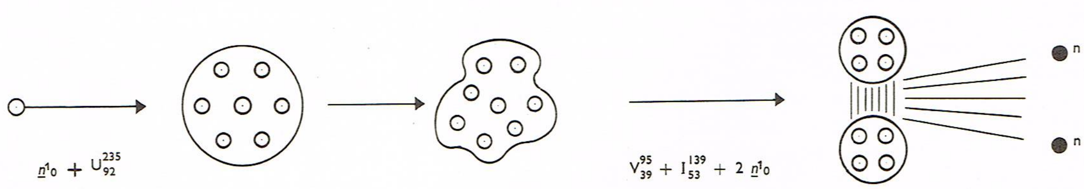
On découvrit d'abord que, des deux isotopes que contenait l'uranium naturel, U^{235} et U^{238}, le premier, représentant 0,7 p. 100 de sa masse, était beaucoup plus facilement « fissile » que le second. La fission de l'uranium 235 augmentait également de façon notable sous l'action de neutrons lents, plus précisément de neutrons thermiques, ainsi appelés en raison de leur vitesse (celle-ci correspond à celle de l'agitation incessante des molécules d'un gaz, dont la température n'est que la traduction). Une autre découverte, due à F. Joliot et à son groupe, devait fournir la clé de la propagation de la fission de proche en proche et, partant, d'une libération intensive d'énergie. On trouva, en effet, que si la fission de l'uranium par des neutrons donnait naissance à deux noyaux compacts, elle s'accompagnait également de l'émission d'autres neutrons que les noyaux lourds contiennent justement en surabondance; ainsi la fission d'un noyau d'U^{235} en libère en moyenne 2,5. Ces neutrons peuvent évidemment à leur tour devenir des agents de fission; dans les conditions habituelles, ils ne produisent que rarement d'autres réactions et la fission ne se continue pas. Si l'on réussit à augmenter suffisamment leur activité, la fission se propage de proche en proche et devient une réaction en chaîne.

Dans une pile atomique, la fission nucléaire est contrôlée et entretenue à une allure assez lente; dans une bombe atomique, elle est ultra-rapide et explosive.

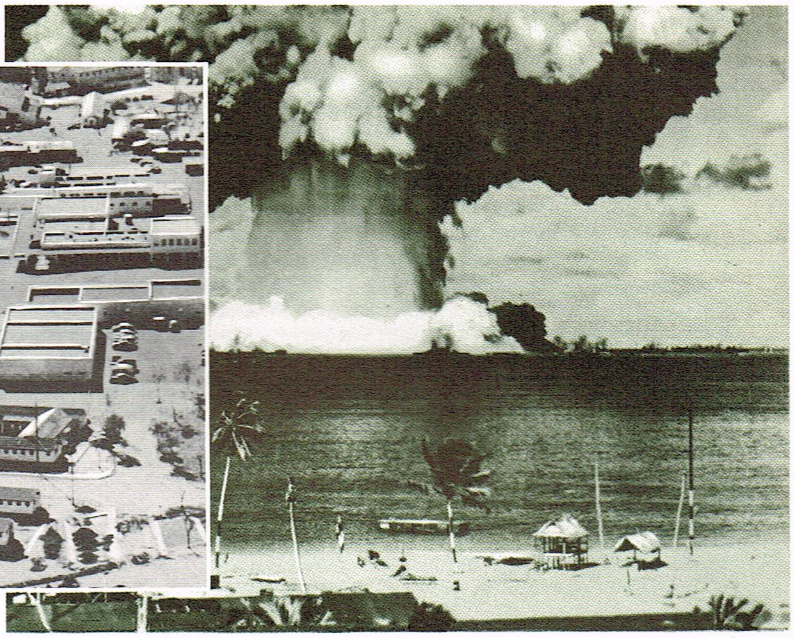
piles et bombes atomiques

La première pile de Fermi était un simple empilement de barres d'uranium naturel et de barres de graphite alternées. La matière première fissile était, bien entendu, l'uranium; comme la réaction en chaîne pouvait difficilement s'établir parce que l'isotope 238 subit malaisément la fission et que l'isotope 235 n'est hautement fissile qu'avec des neutrons lents, des barres de graphite étaient intercalées entre celles d'uranium pour ralentir les neutrons libérés et les rendre plus aptes à la production de nouvelles fissions. L'efficacité des neutrons dépend évidemment de la quantité d'uranium qu'ils traversent, et il avait été nécessaire d'ajuster celle-ci de façon à obtenir une **réaction en chaîne.** Comme il fallait éviter à tout prix que la réaction ne s'emballât et ne devînt explosive sans la laisser pour autant s'arrêter, des trous étaient ménagés dans la masse active pour

Type de fission donnant naissance à des neutrons libres.







Los Alamos, centre où fut construite et mise au point la première bombe atomique. Vivant comme des reclus, soumis au secret absolu, des hommes travaillaient...
Au moment voulu, un détonateur met le feu à la poudre de canon, qui lance violemment les masses d'uranium l'une contre l'autre.

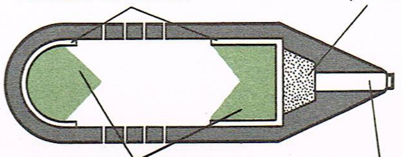


Schéma possible d'une bombe atomique « A ».

pouvoir y glisser des barres d'un corps absorbeur de neutrons. L'intensité du flux de neutrons était mesurée à chaque instant, et les barres de contrôle permettaient de l'ajuster automatiquement au taux voulu.

Toutes les piles sont bâties sur ce principe. La matière première est de l'**uranium naturel** ou de l'**uranium enrichi** en U^{235}. Le ralentisseur ou modérateur, que ce soit le graphite, l'eau lourde, l'eau ordinaire, ou encore le béryllium, est choisi parmi les corps contenant des noyaux légers; il est facile d'imaginer, en effet, que les neutrons ne font que ricocher sur des noyaux lourds, sans être ralentis, tandis qu'ils cèdent une partie de leur vitesse et de leur énergie à d'autres plus légers. Le deutérium, contenu dans l'eau lourde, convient mieux que l'hydrogène, contenu dans l'eau ordinaire, car le proton a tendance à absorber le neutron pour donner justement du deutérium. L'eau lourde est employée avec l'uranium naturel et l'eau ordinaire avec l'uranium enrichi seulement.

Il n'est pas nécessaire ici d'insister sur le rôle des piles atomiques en tant que productrices d'énergie; mais, outre cette libération contrôlée de l'énergie nucléaire, les piles atomiques ont servi à obtenir des flux de neutrons intenses, grâce auxquels des réactions nucléaires ont pu être étudiées et des radio-isotopes (tel le cobalt Co^{60} employé en médecine) produits de façon efficace. Elles ont permis aussi la création des transuraniens, à côté des éléments de fission, et ce fut là une confirmation des premières expériences de Fermi de 1934, qui ne lui avaient donné que des résultats embryonnaires.

liste des transuraniens

nombre atomique	nom	date de première mise en évidence
93	neptunium	1940
94	plutonium	1940
95	américium	1944
96	curium	1944
97	berkélium	1949
98	californium	1950
99	einsteinium	1953
100	fermium	1954
101	mendélévium	1955
102	nobélium	1957

On découvrit en particulier que l'uranium 238, s'il peut subir la fission, donne pourtant plus facilement naissance à un nouvel élément transuranien, le plutonium, dont tout le monde connaît maintenant l'existence. Le neutron s'accole d'abord à l'uranium 238 pour donner un isotope plus lourd, U^{239} :

$$U_{92}^{238} + n_0^1 \rightarrow U_{92}^{239}.$$

Cet isotope U^{239} est instable et se désintègre avec une période courte (23 minutes) en donnant du neptunium, le premier transuranien, suivant le schéma

$$U_{92}^{239} \rightarrow e^- + Np_{93}^{239}.$$

Le neptunium se désintègre à son tour assez vite (période de deux à trois jours) et donne naissance au plutonium :

$$Np_{93}^{239} \rightarrow e^- + Pu_{94}^{239}.$$

Le plutonium est radio-actif, lui aussi, mais la longueur de sa période (24 000 ans) permet de l'isoler aisément. Son importance pratique vient du fait que, comme l'uranium 235, il est très facilement fissile et que cette qualité l'a fait employer dans les bombes atomiques.

C'est, en effet, une matière hautement fissile — uranium 235 ou plutonium — qui constitue la matière première de la bombe atomique « ordinaire ». On conçoit que la fission devienne explosive à partir d'une certaine masse — appelée **masse critique.** En effet, la fission d'un noyau d'uranium 235, par exemple, libère en moyenne 2,5 neutrons. Si ces neutrons se propagent dans une masse suffisante d'uranium 235 intact, ils vont pouvoir causer d'autres fissions, qui libéreront d'autres neutrons, et ainsi neutrons libres et fissions vont augmenter de pair. Ces neutrons sont rapides et mettent à peu près 10^{-5} seconde à produire une autre fission ou à sortir de la matière fissile. Il est clair qu'avec une masse suffisante d'élément fissile, la réaction sera explosive. Au contraire, en dessous de cette masse critique les neutrons s'échapperont trop souvent et la fission s'arrêtera toute seule. **Le principe de fonctionnement d'une bombe atomique consiste donc à rassembler rapidement deux masses sous-critiques de façon à obtenir une masse sur-critique.** La réaction en chaîne commence alors, se propage à une vitesse vertigineuse, et la bombe

Phot. U. S. I. S.

explose, libérant en quelques secondes une énergie fantastique. La masse critique de l'uranium 235 est environ de 14 kg, celle du plutonium de 5. Construire une bombe n'est pas une entreprise facile, car, pour produire 1 kg de plutonium, une pile atomique de 1 000 kW demande 1 000 jours, et, pour isoler 10 kg d'uranium 235, il faut traiter plus d'une tonne et demie d'uranium naturel. La séparation des deux isotopes U^{235} et U^{238} de l'uranium est, de plus, un problème très difficile. Aucune méthode chimique n'est évidemment possible, puisque, comme tous les isotopes, ces deux corps ont exactement les mêmes propriétés chimiques. Il faut recourir à des méthodes fondées sur la seule différence qu'ils présentent : leur différence de masse. Mais celle-ci est faible, et la séparation n'en est que plus ardue. L'un des procédés employés consiste à faire diffuser, à travers des parois poreuses, un composé gazeux d'uranium; l'uranium 235, un peu plus léger que l'uranium 238, diffuse un peu plus rapidement, et, moyennant de nombreuses diffusions successives, le mélange peut être enrichi en uranium 235. Comme, de plus, le composé gazeux employé est très corrosif, les problèmes techniques se compliquent encore, et c'est la construction d'une usine extrêmement coûteuse qu'il faut entreprendre.

Ces bombes atomiques, à l'uranium 235 ou au plutonium, ont des puissances limitées, puisqu'on ne peut se permettre de dépasser la masse critique pour un morceau compact d'élément fissile. Les bombes thermonucléaires, au contraire, sont plus puissantes et ne connaissent pas cet inconvénient, si inconvénient il y a! Autour d'une bombe ordinaire est disposé un corps constitué de noyaux légers capables de subir la fusion. L'élévation de température — quelques millions de degrés — produite par la fission de la bombe interne provoque la fusion des noyaux légers et, partant, dégage une quantité supplémentaire d'énergie. Il n'y a pas de problème de masse critique pour les éléments de la fusion, et la puissance des bombes n'est plus limitée que par leur encombrement.

Mais, pour bien comprendre pourquoi c'est l'élévation de température d'une bombe atomique qui provoque la fusion des noyaux légers, il est nécessaire de s'étendre un peu plus sur ces questions.

la fusion nucléaire et ses problèmes

Les réactions de fusion sont celles où deux noyaux se mêlent intimement l'un à l'autre au point de perdre leur individualité et donnent naissance à un nouveau noyau plus lourd. Du point de vue énergétique qui nous occupe ici, la fusion n'a évidemment d'intérêt que lorsqu'elle s'accompagne d'une libération d'énergie due à la plus grande stabilité des nouveaux noyaux formés.

Les possibilités théoriques sont nombreuses; il suffit de regarder le tableau des énergies de liaison des nucléons, différentes suivant les éléments, pour imaginer nombre de fusions intéressantes. Par la fusion, on pourrait ainsi fondre deux noyaux d'oxygène O_8^{16} en un noyau de soufre S_{16}^{32} et récolter une énergie appréciable de 18 MeV.

En fait, l'exploitation énergétique de la fusion nucléaire n'est pas encore une réalité, et il nous faut bien en comprendre les raisons.

Remarquons tout d'abord que le neutron, le meneur de jeu de la fission, n'est ici d'aucun secours. La fission d'un noyau lourd par un neutron s'accompagne de l'émission de plusieurs autres neutrons et la réaction peut s'entretenir; la fusion d'un neutron avec un noyau « léger » n'en libère aucun autre et aucune réaction en chaîne n'est ainsi possible.

Force est donc de s'adresser à des noyaux chargés, existant d'ailleurs en abondance dans la nature, et de trouver un moyen de les faire fusionner avec un rendement suffisant. Produire des réactions nucléaires entre noyaux chargés n'est cependant pas un problème simple. Une première difficulté nous est déjà connue : pour amener deux noyaux au contact l'un de l'autre, prémisse indispensable à leur fusion, il faut leur donner une énergie cinétique suffisante pour vaincre les forces de répulsion qui s'exercent entre les charges électriques qu'on veut ainsi rapprocher. Ces forces entre charges varient comme l'inverse du carré de leur distance et deviennent très importantes quand les distances d'approche sont aussi petites que celles du rayon nucléaire. Par ailleurs, plus lourd est le noyau, plus nombreuses sont les charges électriques et plus intenses les forces de répulsion correspondantes. C'est bien pour cette raison que le difficile problème de la fusion se restreint en pratique à la seule fusion des éléments légers, hydrogène, deutérium et tritium, qui, tous, ont un noyau contenant un seul proton.

L'énergie nécessaire pour vaincre la répulsion électrique entre deux protons que l'on désire accoler est de 0,5 MeV. C'est une énergie importante, puisque l'on sait qu'il faut accélérer un proton sous une différence de potentiel de 1 million de volts pour lui fournir une énergie de 1 MeV. Cette difficulté a été résolue et on a réussi à obtenir des réactions de fusion; on a construit des accélérateurs permettant de soumettre les particules à des différences de potentiel de quelques millions de volts; on a découvert aussi que les particules arrivant devant la barrière électrique du noyau parvenaient de temps en temps à se faufiler, même si elles n'avaient pas l'énergie suffisante pour « sauter » la barrière : ainsi ce n'est pas à partir de 0,5 MeV, mais de 0,01 MeV qu'un proton peut fusionner avec un autre proton. Cet « effet tunnel » (dû à la nature ondulatoire de la lumière) rend évidemment plus simples déjà les possibilités de fusion. En envoyant des protons, deutons ou tritons accélérés sur l'hydrogène, le deutérium ou le tritium, on a ainsi pu observer les diverses réactions de fusion.

Malheureusement, le rendement ainsi obtenu est dérisoire. Une particule accélérée n'a qu'une chance sur dix mille environ de produire une fusion; l'énergie dépensée dans l'accélération des particules est plus grande que celle qu'il est possible de récupérer. La raison de l'échec est la suivante : les particules accélérées arrivant sur la cible dépensent leur énergie à chasser les électrons beaucoup plus facilement qu'à produire une fusion. Rapidement ralenties, elles deviennent inaptes à la fusion et finissent simplement par s'arrêter; le parcours d'un proton de 2 MeV dans une feuille d'aluminium est ainsi de 0,01 mm. Certes, si on leur donne une énergie plus grande, les particules traversent plus de matière, mais leur probabilité de rencontre avec les noyaux n'augmente pas sensiblement, car, plus rapides, elles passent très vite près d'un noyau et sont moins tentées de

$$H_1^1 + H_1^1 \rightarrow H_1^2 + e^+ + 1{,}7\ \text{MeV}$$
$$H_1^1 + H_1^2 \rightarrow H_2^3 + 5\ \text{MeV}$$
$$H_1^1 + H_1^3 \rightarrow H_2^4 + 19\ \text{MeV}^*$$
$$H_1^2 + H_1^2 \rightarrow \begin{cases} H_2^3 + n_0^1 + 3{,}5\ \text{MeV} \\ H_1^3 + H_1^1 + 4\ \text{MeV} \end{cases}$$
$$H_1^2 + H_1^3 \rightarrow H_2^4 + n_0^1 + 17{,}6\ \text{MeV}^*$$
$$H_1^3 + H_1^3 \rightarrow H_2^4 + 2n_0^1 + 11{,}4\ \text{MeV}^*$$

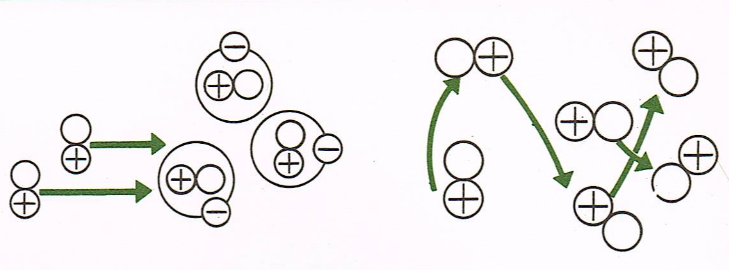
Réactions de fusion des noyaux légers. Les réactions libérant le plus d'énergie sont celles qui donnent naissance au noyau d'H_4^e, exceptionnellement stable (* dans le tableau). Dans la bombe à hydrogène, les réactions utilisées sont les réactions *pd* (proton + deuton) et *pt* (proton + triton); dans les essais de fusion contrôlée, on s'adresse aux réactions *dd* et *dt*; dans le Soleil, on a la succession des deux premières réactions *pp* puis *pd*, suivies de la fusion $H_2^3 + H_2^3 \rightarrow H_2^4 + 2H_1^1$.

s'arrêter. **Aucun espoir ne subsiste d'avoir par cette méthode une libération rentable d'énergie de fusion.** Pour sortir de l'impasse, il faudrait envoyer les particules accélérées sur une cible constituée d'atomes ionisés, c'est-à-dire débarrassés de leurs électrons, mais alors les données du problème changeraient complètement. En effet, pour ioniser tout un ensemble d'atomes, il faut donner aux électrons et aux noyaux des vitesses suffisantes pour les rendre libres et empêcher leur recombinaison; ces vitesses donnent aux lourds noyaux des énergies cinétiques qui leur suffisent déjà pour se heurter de plein fouet et fusionner, au hasard de leurs déplacements. Il n'est plus besoin d'employer un accélérateur et d'envoyer les particules sur une cible; c'est le milieu même de la cible qui entre en effervescence, et ainsi c'est vers la constitution d'un mélange de noyaux et d'électrons complètement libres, qu'on appelle encore un « plasma », qu'il faut se tourner pour trouver la bonne solution.

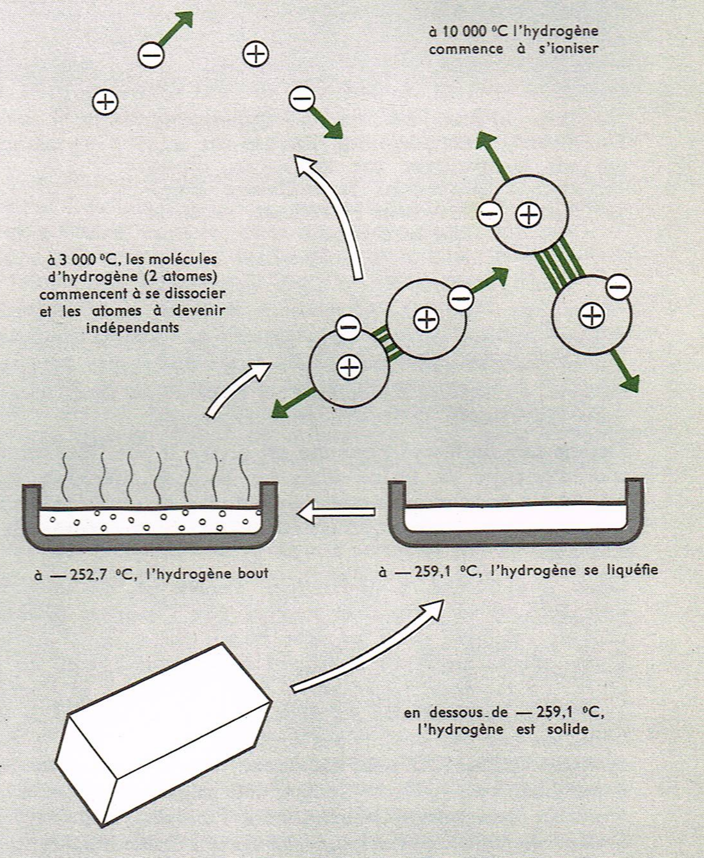
bombe à hydrogène et Soleil

Il y a une relation étroite entre la température d'un corps et l'agitation continuelle de ses molécules ou atomes. A basse température, un corps est solide; les molécules sont fortement attachées les unes aux autres et leur léger mouvement n'entame pas la cohésion de l'ensemble. Que se passe-t-il quand on chauffe ce corps? L'énergie calorifique apportée sert à augmenter l'énergie cinétique des molécules, qui s'agitent de plus en plus, dans un mouvement désordonné. La belle cohésion du solide s'en trouve compromise, les liens deviennent plus lâches, le solide se transforme en liquide. Si l'on chauffe encore, l'agitation croît de plus belle et les molécules deviennent capables de s'échapper dans l'atmosphère : le liquide bout. Le volume d'un gaz ne connaît alors d'autres limites que celles de l'enceinte qui le contient, car les molécules, infatigables, tentent toujours de gagner le large. Si elles rencontrent une paroi solide, elles sont obligées de rebondir et restent emprisonnées : ces chocs nombreux sont la cause de la pression exercée par le gaz sur les parois. A plus haute température encore, les molécules du gaz, constituées en général de deux atomes, se dissocient à leur tour, et les atomes deviennent indépendants. C'est à 10 000 °C environ que commence l'ionisation de l'atome, et il faut atteindre des températures de quelques millions de degrés pour qu'il y ait un plasma, gaz complètement ionisé, capable de subir la fusion.

Ainsi la fusion nucléaire est liée à la réalisation de très hautes températures. Le Soleil et les étoiles sont justement des foyers où règnent des conditions qui rendent possibles les réactions thermonucléaires de fusion. Des calculs ont montré que si le Soleil était une immense chaudière à charbon, il n'aurait pu rayonner pendant plus de cinq siècles et aurait dû être consumé depuis longtemps. Il existe des raisons pour

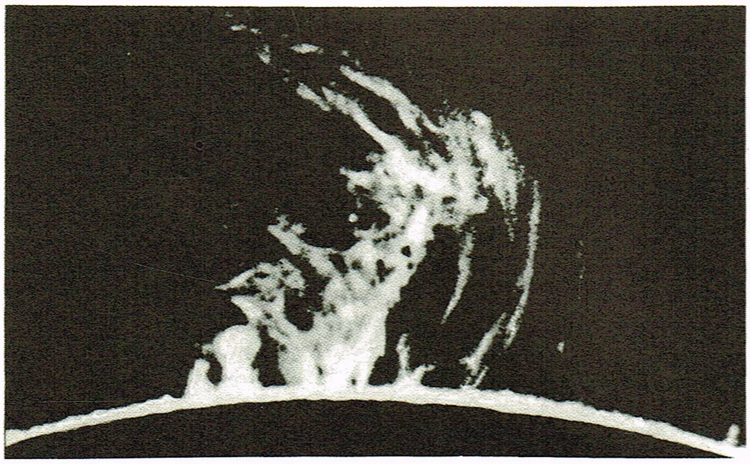


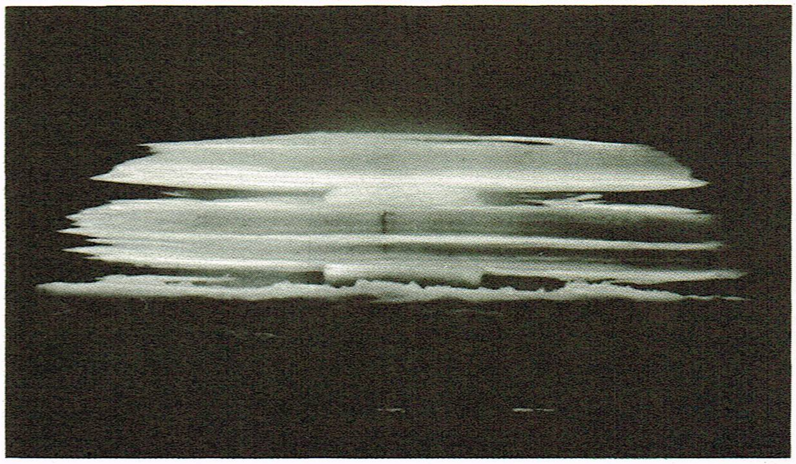
Deux méthodes pour obtenir la fusion de noyaux légers : la première — deutons accélérés envoyés sur du deutérium — a un rendement infime ; la seconde seule — deutons cognant d'autres deutons — serait prometteuse, mais se heurte à plus d'une difficulté.



États physiques de l'hydrogène suivant la température (à pression atmosphérique).

Soleil et bombe « H » empruntent leur énergie à la fusion.





Doc. Observatoire du Mont-Wilson et phot. Associated Press.

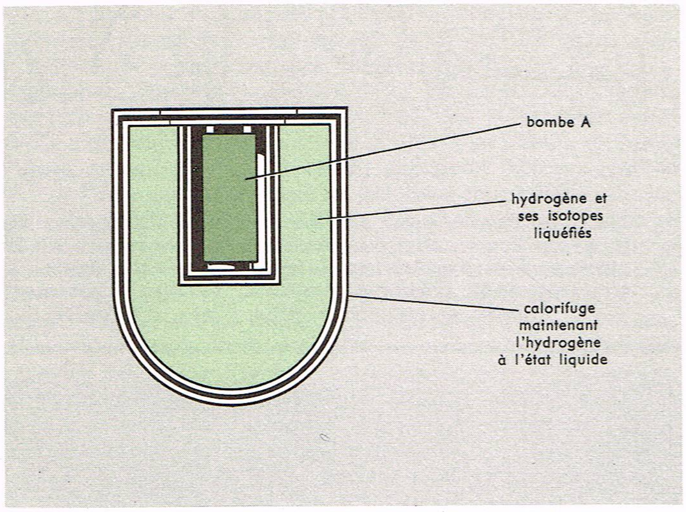


Schéma d'une bombe à hydrogène. La bombe classique à l'uranium 235 ou au plutonium est entourée d'une enveloppe contenant le mélange léger tel que deutérium + tritium, source de la fusion.

penser que son énergie vient surtout de noyaux d'hydrogène, qui se transforment peu à peu, par différentes fusions successives, en noyaux d'hélium, dont la présence assez abondante dans notre astre se trouve ainsi expliquée.

Depuis peu de temps l'homme est arrivé à une libération massive d'énergie de fusion dans les bombes thermo-nucléaires mais il n'a pas encore réussi pour autant à résoudre le problème beaucoup plus compliqué d'une libération **contrôlée** de cette énergie.

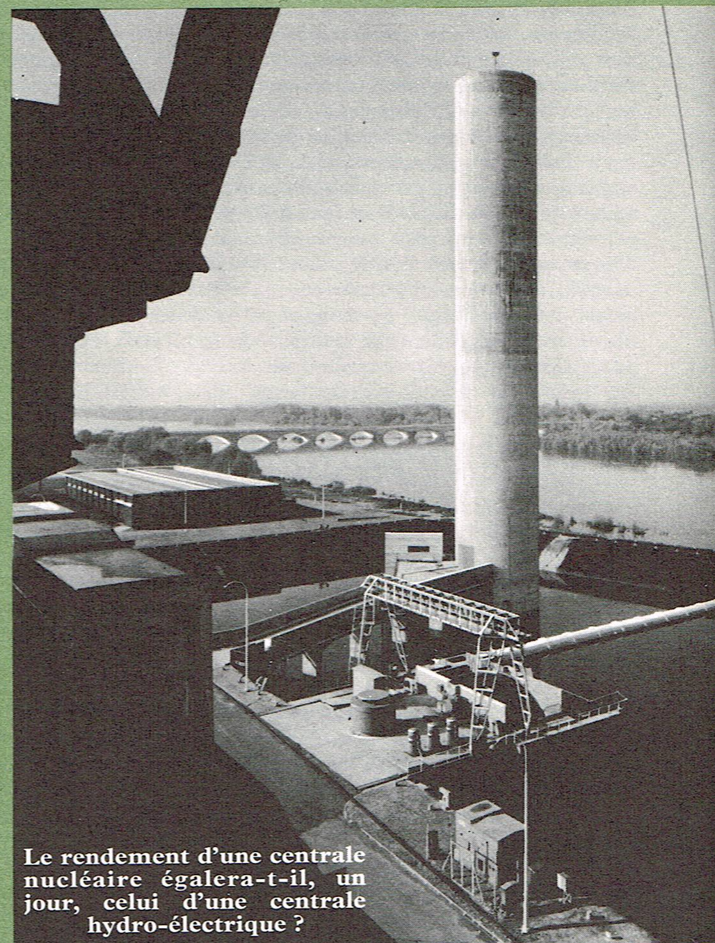
Dans une bombe à hydrogène, l'élévation de température nécessaire à la fusion est fournie par l'énorme quantité de chaleur dégagée dans un très court temps par l'éclatement d'une bombe atomique ordinaire à uranium 235 ou au plutonium. Le mélange léger entre alors en fusion et libère une quantité supplémentaire d'énergie, qui ne connaît pas de limites imposées par des dimensions critiques, comme dans le cas des matière fissiles. On peut fabriquer des bombes à hydrogène encombrantes peut-être, mais d'une puissance quasi illimitée. On s'est évidemment tourné vers l'utilisation d'éléments de fusion donnant le meilleur rendement possible, et adressé pour cela aux mélanges hydrogène-tritium ou deutérium-tritium; le tritium est d'ailleurs cher, difficile à obtenir et à conserver, puisqu'il est radio-actif. Il est maintenant fabriqué sur place grâce à l'emploi du lithium, sous forme d'hydrure de lithium en particulier : l'éclatement de la bombe atomique et les neutrons qu'elle projette de toutes parts permettent d'abord de transformer le lithium en tritium suivant la réaction

$$Li_3^6 + n_0^1 \longrightarrow H_1^3 + H_2^4.$$

Ensuite le tritium et l'hydrogène fusionnent. La bombe H a fonctionné.

vers la fusion contrôlée

Comment obtenir en laboratoire les millions de degrés nécessaires à la fusion? A ces températures, aucune paroi solide ne résiste, et une paroi solide est pourtant nécessaire pour ne pas laisser s'échapper le plasma. De plus, l'isolement est indispensable si l'on ne veut pas que l'énergie calorifique apportée se perde par conduction thermique au lieu de chauffer le plasma. Ces problèmes, qui peuvent sembler insolubles à première vue, ont pourtant reçu un début de solution. La bombe H a fonctionné.



Le rendement d'une centrale nucléaire égalera-t-il, un jour, celui d'une centrale hydro-électrique ?

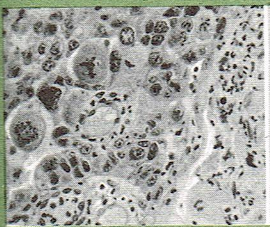


Grâce à l'énergie nucléaire, le brise-glace *Lénine* peut rester longtemps loin de toute base.

Plusieurs méthodes ont été avancées pour confiner le plasma à l'intérieur de son enveloppe sans qu'il vienne la toucher.

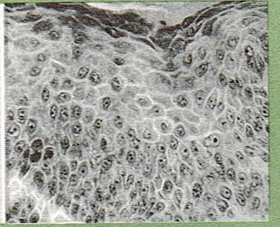
La plus originale est peut-être celle qui a été essayée en Angleterre dans l'appareil « Zéta », qui a beaucoup fait parler de lui, il y a quelques années, avec l'annonce, prématurée et fausse, de la réussite de la fusion. Zéta est constitué par un anneau de 3 m de diamètre environ, une espèce de couronne à l'intérieur de laquelle se trouve le mélange gazeux léger. Le mélange est chauffé par des décharges électriques à l'intérieur du tube et s'ionise pour devenir un plasma. Il se passe alors une chose curieuse et très avantageuse : soumis à un effet de striction, le plasma se resserre à l'intérieur de l'anneau et se détache

Phot. Jahan (doc. C. E. A.), Bureau soviétique d'inform.



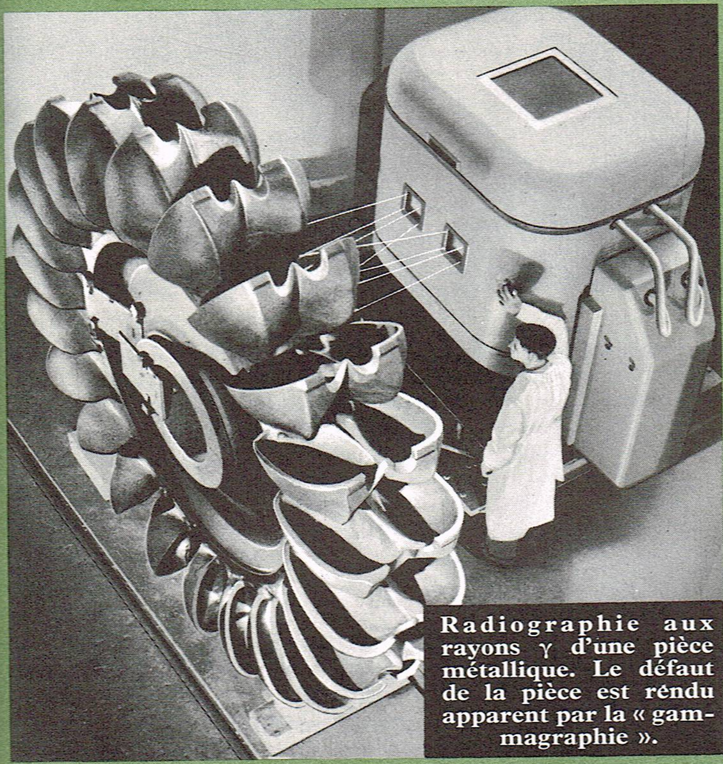
◀ Tissu cancéreux et tissu normal. ▶

La bombe au cobalt rendra peut-être sain le tissu malade.





M. Khrouchtchev exhibant une variété de maïs au rendement exceptionnel : les rayons γ sont-ils passés par là ?



Radiographie aux rayons γ d'une pièce métallique. Le défaut de la pièce est rendu apparent par la « gammagraphie ».

des parois. La raison en est que les particules électrisées, en se déplaçant sous l'effet de la décharge, créent un champ magnétique qui les attire tout comme le font entre eux deux fils électriques parcourus par des courants de même sens. Ainsi, des décharges électriques très courtes et très intenses (des courants de quelques dizaines de milliers d'ampères en quelques microsecondes, obtenus par des décharges de condensateurs) permettent à la fois le chauffage du plasma par l'effet de Joule et son confinement par effet de striction.

Malheureusement, ce confinement est instable et le plasma se déforme très vite. Il vient heurter les parois, et la très haute température ne peut être maintenue que pendant quelques millisecondes. Là est l'une des pierres d'achoppement de cette fusion contrôlée. Les recherches marquent actuellement le pas, mais différentes équipes, aux Etats-Unis, en U. R. S. S., en Grande-Bretagne et en France, travaillent dans le silence, et le jour de la réussite n'est peut-être pas tellement éloigné.

***

Qui dira l'avenir de l'énergie nucléaire dans notre monde? L'homme est encore ahuri devant la puissance de l'énergie qu'il a réussi à dompter, puissance qui pourrait finalement se retourner contre lui et l'anéantir à tout jamais. La pensée d'une telle éventualité nous remplit d'effroi, mais, indépendamment de tout emploi tragique, l'utilisation d'une telle puissance fait toujours figure d'un risque aux conséquences difficiles à prévoir. Qui donc se fait fort de connaître à long terme toutes les conséquences héréditaires des radiations nucléaires? **La maîtrise de toute technique moderne devient de plus en plus difficile, et ici les implications sont aux dimensions mêmes du monde et rendent les hommes dépendants tragiquement les uns des autres. Qu'on le veuille ou non, le dialogue et même la collaboration — sinon la bonne entente — s'imposeront inéluctablement.**

Quel bien peut-on donc attendre de cette puissance de l'énergie nucléaire? Il serait facile d'ironiser et d'inventorier ici tout un arsenal militaire, mais les piles atomiques ne servent pas qu'à produire du plutonium et des bombes atomiques; elles peuvent nous fournir de l'électricité, avec un rendement, il est vrai, inférieur pour le moment à celui des centrales hydro-électriques. Mais demain la situation sera peut-être complètement bouleversée avec l'apparition de l'énergie de fusion contrôlée. Le problème de la propulsion nucléaire commence à faire des progrès; elle est déjà utilisée sur des navires, auxquels elle donne une grande liberté de manœuvre. Qui ne se rappelle l'exploit accompli par le sous-marin américain *Skate* naviguant sous les glaces polaires?

Pour n'être pas aussi impressionnantes, d'autres applications n'en rendent pas moins de précieux services. Les effets sur les cellules vivantes des radiations nucléaires, et plus précisément des rayons γ, frères des rayons X mais plus énergiques, sont encore loin d'être tous mis en valeur. Des tumeurs cancéreuses sont traitées à la « bombe au cobalt », des affections de la glande thyroïde à l'iode radio-actif, et les rayons γ commencent à être utilisés dans les techniques de stérilisation et de reproduction sélective des espèces végétales.

Dans beaucoup de domaines, la radio-activité offre des solutions élégantes et faciles par l'entremise de la technique générale dite des « noyaux traceurs ». Supposons, par exemple, que l'on veuille connaître le trajet et la localisation d'aliments dans le corps humain. On peut y arriver en ajoutant aux composés contenant du carbone ordinaire une faible proportion de ces mêmes composés, identiques du point de vue chimique et biologique, mais contenant du carbone 14 radio-actif. Ce dernier suit fidèlement les aliments dans leur acheminement et, à tout moment, il peut jouer le rôle d'indicateur grâce à sa radio-activité facilement détectable. Ainsi cette technique rend de précieux services lorsqu'il s'agit de suivre des phénomènes difficilement observables. Elle est employée dans des tâches aussi diverses que la localisation de tumeurs cérébrales ou que le contrôle du passage de différentes qualités de pétrole se suivant sans discontinuité dans un pipe-line.

Si l'on se souvient aussi des possibilités de datation des minéraux et des végétaux par l'uranium et le carbone 14, dont nous avons parlé assez longuement, on voit que longue est la liste des applications des radiations nucléaires, qui viennent augmenter les pouvoirs de la médecine et de la biologie et enrichir l'agriculture, la géologie, l'archéologie...

Mais cela ne suffirait sans doute pas à nous satisfaire si nous ne gardions l'espoir que le fléau de la bombe atomique ne force les hommes à s'entendre.

Phot. Keystone, Brown-Boveri.

physique nucléaire des hautes énergies

source et détection des particules de haute énergie

Quand on connaît la puissance terrifiante d'une bombe atomique, il peut sembler paradoxal d'entendre dire que fission ou fusion relèvent de la physique nucléaire de « basse » énergie. Il ne faut surtout pas se méprendre : les bombes H gardent toujours, et de loin, le monopole d'une libération intensive d'énergie; la différence entre « basse » et « haute » énergie n'est pas à estimer à notre niveau à nous, celui des applications pratiques, mais à celui du noyau.

Jusqu'ici les phénomènes nucléaires dont nous avons parlé — fission et fusion y compris — mettaient en jeu pour les nucléons des énergies de l'ordre de grandeur de celles qui les lient entre eux à l'intérieur d'un noyau. On étudie maintenant des réactions où les énergies considérées sont beaucoup plus importantes que celles qui président à cet arrangement des nucléons dans la nature. La physique nucléaire des basses énergies s'intéresse en premier chef aux transformations des noyaux entre eux et l'énergie cinétique des particules étudiées est de quelques dizaines de méga-électron-volts tout au plus. Pour des énergies plus grandes, disons à partir d'une centaine de méga-électron-volts, bien qu'il n'y ait pas de frontière définie, s'ouvre la physique des « hautes énergies ».

L'exploration de ce nouveau domaine a **commencé par la découverte et l'étude des rayons cosmiques,** gracieusement envoyés par le ciel, tout comme la découverte et l'étude de la radio-activité naturelle avaient marqué le point de départ de la physique nucléaire en général. Dans un cas comme dans l'autre, grâce aux accélérateurs, **la production artificielle des particules énergiques fut l'occasion d'un véritable bond en avant dans le développement de nos connaissances.**

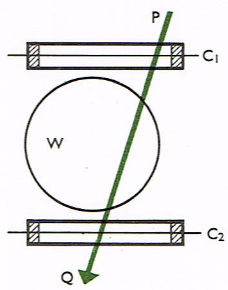
Avant d'essayer de résumer les connaissances que nous avons des particules que l'on se plaît à appeler « élémentaires », nous allons passer successivement en revue les moyens de créer et de détecter ces particules énergiques.

une source naturelle, les rayons cosmiques

L'histoire des rayons cosmiques commença vers 1900, quand on s'aperçut qu'une chambre à ionisation ne cessait jamais d'être excitée, même si elle était bien isolée de toute source extérieure. La radio-activité naturelle, découverte depuis peu, était encore mal connue, et ce « bruit de fond » fut d'abord mis au compte de quelque radio-activité inconnue venant du sol. Les premières expériences faites en ballon vers 1910 montrèrent, au contraire, que le bruit de fond trouvé au sol augmentait en altitude; il était donc dû à un rayonnement extra-terrestre et il reçut le nom de rayonnement « cosmique » quand on s'aperçut qu'il était aussi intense la nuit que le jour et qu'apparemment il ne dépendait ni de la position du Soleil, ni de celle de la Lune, ni même de celle de la Voie lactée.

L'étude de ce nouveau rayonnement n'était pas facile, puisqu'il avait une intensité très faible. Nous savons maintenant que sur une surface de 1 cm^2 au niveau de la mer, il arrive en moyenne un rayon cosmique par minute. C'est peu si nous comparons ce nombre aux millions de millions d'électrons qui constituent un courant de 1 μA! En revanche, ces particules peuvent avoir individuellement de très grandes énergies, comme le montrèrent de premières expériences d'absorption par la matière. Si une chambre d'ionisation est entourée par des épaisseurs grandissantes de plomb, le rayonnement de la radio-activité naturelle peut être arrêté, le rayonnement cosmique est certes diminué, mais il n'est jamais complètement supprimé.

Les physiciens s'adonnèrent donc d'abord à des études d'absorption, méthode habituellement employée pour commencer à déchiffrer un rayonnement inconnu, mais la nature de ce dernier semblait complexe et ils ne pouvaient véritablement progresser qu'en utilisant d'autres moyens de détection que la simple chambre à ionisation, autres moyens que réclamait aussi, dans le même temps, l'étude de la radio-activité naturelle. La **chambre de Wilson**, qui vit le jour en 1911 et qui permettait de voir les trajectoires des particules chargées, une à une, devait être l'un de ces instruments de choix à partir du moment où la mise au point de compteurs Geiger permit son déclenchement lors du passage certain de particules. Il y avait ainsi possibilité de saisir au vol les rares rayons cosmiques et leurs interactions, plus rares encore.



Technique de déclenchement automatique d'une chambre de Wilson (W) sur passage de rayons cosmiques de P à Q. Les compteurs Geiger (C_1 et C_2) enregistrent le passage de particules individuelles et commandent électroniquement le fonctionnement de la chambre de Wilson. La particule a disparu, mais a laissé derrière elle des ions qui donnent naissance aux gouttelettes liquides. Différents arrangements de compteurs sont possibles; si l'on veut, par exemple, que les particules traversent la chambre de part en part, on peut disposer une rangée de compteurs en haut, une autre en bas, et ne faire fonctionner la chambre que si deux compteurs (un de chaque rangée) donnent un signal en « coïncidence ».

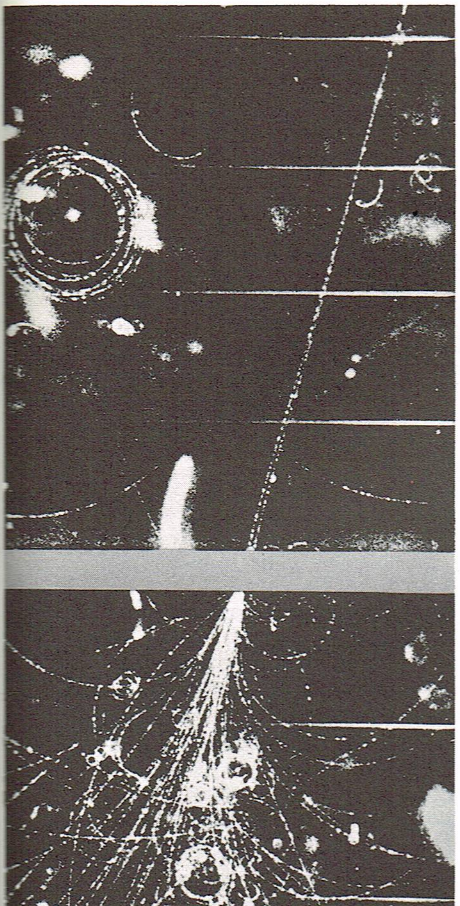
Cet instrument si intéressant, les physiciens le perfectionnèrent encore pour en augmenter les potentialités. Pour mieux étudier le comportement des rayons cosmiques à la traversée de la matière, ils introduisirent à l'intérieur même de la chambre des plaques métalliques. Pour avoir des renseignements sur les particules, sur leur énergie et le signe de leur charge électrique, ils entourèrent l'appareil d'un électro-aimant aussi puissant que possible. En effet, soumise ainsi à un champ magnétique B, la particule décrit un arc de cercle, dont la mesure sur photographie du rayon R permet de déterminer l'impulsion P de la particule par la relation simple

$$P = 300\,B\,R \quad \begin{cases} B \text{ étant en Web/m}^2\ (10^4 \text{ gauss}), \\ R \quad — \quad \text{m}, \\ P \quad — \quad \text{MeV/c}. \end{cases}$$

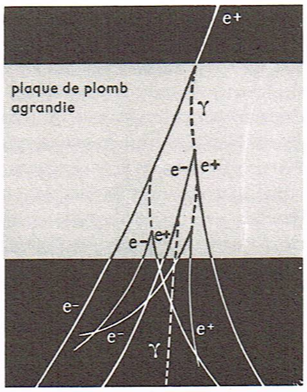
Est-il besoin de dire que cette formule est conforme aux lois de la mécanique relativiste? Ici, en effet, les vitesses des particules se rapprochent de celle de la lumière, et la mécanique classique ne suffit plus. **La mécanique relativiste, à laquelle la physique des hautes énergies n'apporte que des confirmations, devient en ce domaine aussi naturelle que l'était la mécanique classique dans l'étude de la chute des corps par gravité.** Il est bon, avant d'aller plus loin, de se rappeler les définitions et formules de cette nouvelle mécanique, qui se réduiront d'ailleurs à celles de l'ancienne pour des vitesses faibles par rapport à celle de la lumière.

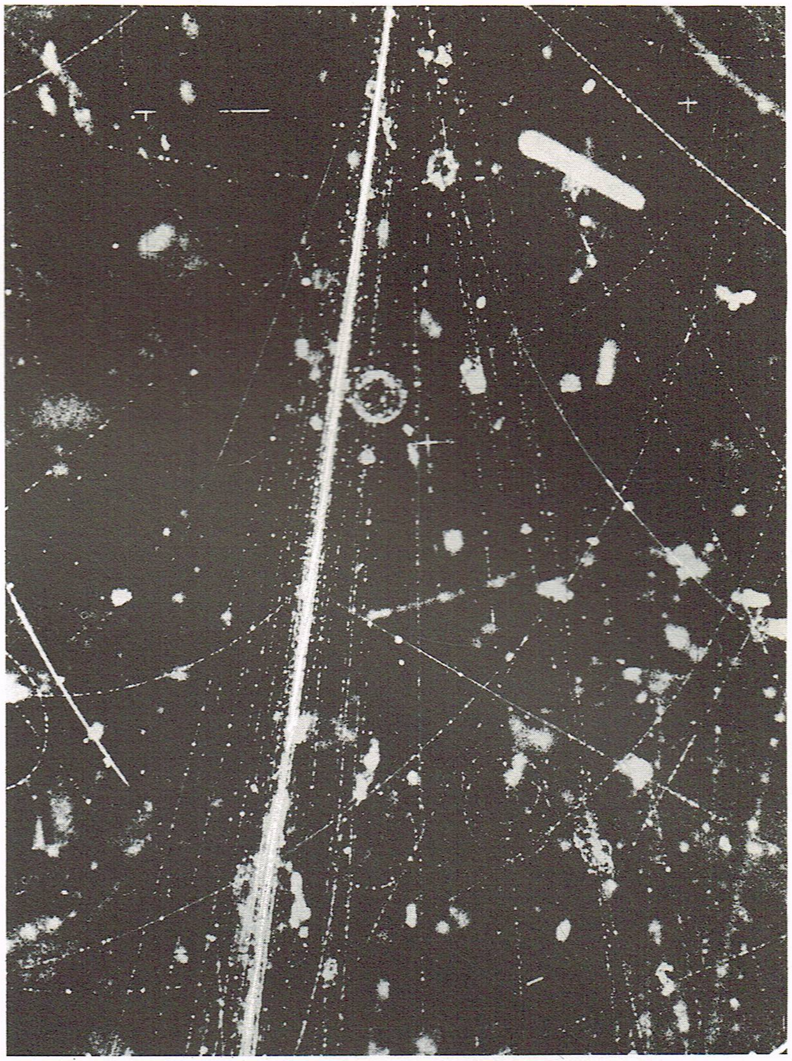
Une particule au repos, de masse *m*, a une énergie interne (dite encore « énergie de masse ») $\mathbf{E_0 = mc^2}$. La masse de la particule est ainsi, en physique des hautes énergies, mesurée conventionnellement comme une énergie, en méga-électron-volts.

Une particule en mouvement, de vitesse *v*, a une énergie totale E, une énergie cinétique T et une impulsion P définies par les formules suivantes :



En traversant la matière (ici une plaque de plomb), les électrons énergiques sont ralentis et émettent des rayons γ qui, eux-mêmes, se convertissent en paires $e^+ + e^-$. Il y a ainsi multiplication d'électrons après traversée de la plaque.





Les rayons cosmiques sont rares, mais leur manifestation est parfois intempestive. Ici, l'un d'eux fait irruption dans la chambre de Wilson et fournit une gerbe de particules rapides.

$$E = \frac{mc^2}{\sqrt{1-\beta^2}}, \quad T = E - mc^2 \ (\beta = v/c),$$

$$P = \frac{mv}{\sqrt{1-\beta^2}} \text{ ou encore } Pc = \sqrt{E^2 - m^2c^4}.$$

Le produit Pc, équivalent à une énergie, peut s'exprimer en MeV, et donc P en MeV/c, comme il a été noté plus haut.

Si la vitesse v est faible devant c ($\beta \ll 1$), le terme $\frac{1}{\sqrt{1-\beta^2}}$ est proche de $1 + {}^1/_2\,\beta^2$ et les formules se réduisent alors par approximation à celles de la mécanique classique :

$$E = mc^2 + \tfrac{1}{2} mv^2, \quad T = \tfrac{1}{2} mv^2, \quad P = mv.$$

La technique perfectionnée de la chambre de Wilson devait porter ses fruits. L'utilisation d'un champ magnétique devait permettre à Anderson, en 1932, d'observer une particule positive qui ne pouvait pas être un proton, parce qu'il aurait dû donner une trace bien plus noire que celle qui était observée. L'**électron positif** était découvert ainsi dans le rayonnement cosmique avant d'être observé un peu plus tard en radio-activité. Peu après, on découvrit que ces e^+ étaient créés par des rayons γ de haute énergie, qui s'annihilaient au contact de la matière et se convertissaient en une paire d'électrons $e^+ + e^-$. Par analogie avec le rayonnement visible ce phénomène montrait avec netteté que la « lumière » se présentait bien ici comme formée de grains individuels bien localisés, non comme une onde envahissant l'espace. Fallait-il revenir sur la vieille querelle : la lumière est-elle une onde ou un corpuscule?

Vers 1932 aussi, une particule nouvelle commença à être soupçonnée qui aurait expliqué que le rayonnement cosmique pût être aussi pénétrant, beaucoup plus pénétrant que celui des protons et aussi que celui qui était formé par ces gerbes d'électrons, nombreuses certes, mais qu'une épaisseur convenable de plomb finissait par arrêter. En 1937, cette particule était découverte et, dans une double chambre de Wilson, on put mesurer sa masse, intermédiaire entre celle de l'électron et celle du proton, raison pour laquelle on l'appela **méson** μ (du grec *mesos*, « moyen »).

Cette découverte marqua le début d'un renouveau pour les rayons cosmiques. Le méson μ pouvait peut-être dévoiler le secret des forces nucléaires; en 1935, en effet, un physicien japonais, Yukawa, avait supposé l'existence d'un « méson » pour expliquer les interactions nucléaires, et voici que le méson μ avait à peu près la masse prévue! Hélas! il y avait une énigme : si le méson μ était si pénétrant, c'est qu'il interagissait fort peu avec la matière. Conduite singulière et, à vrai dire, incompréhensible pour cette particule responsable des interactions nucléaires!

L'hypothèse de Yukawa n'était pas fausse, mais elle ne devait se vérifier qu'en 1947, date à laquelle fut découvert un autre méson qui répondait beaucoup mieux aux prévisions. Cette découverte fut précédée par l'invention d'un nouveau moyen de détection : Powell et Occhialini, en Grande-Bretagne, mirent au point une émulsion photographique spéciale à grains très fins, capable de révéler le passage individuel de particules du rayonnement cosmique. En exposant ces émulsions nucléaires en haute montagne, là où les rayons cosmiques, moins absorbés

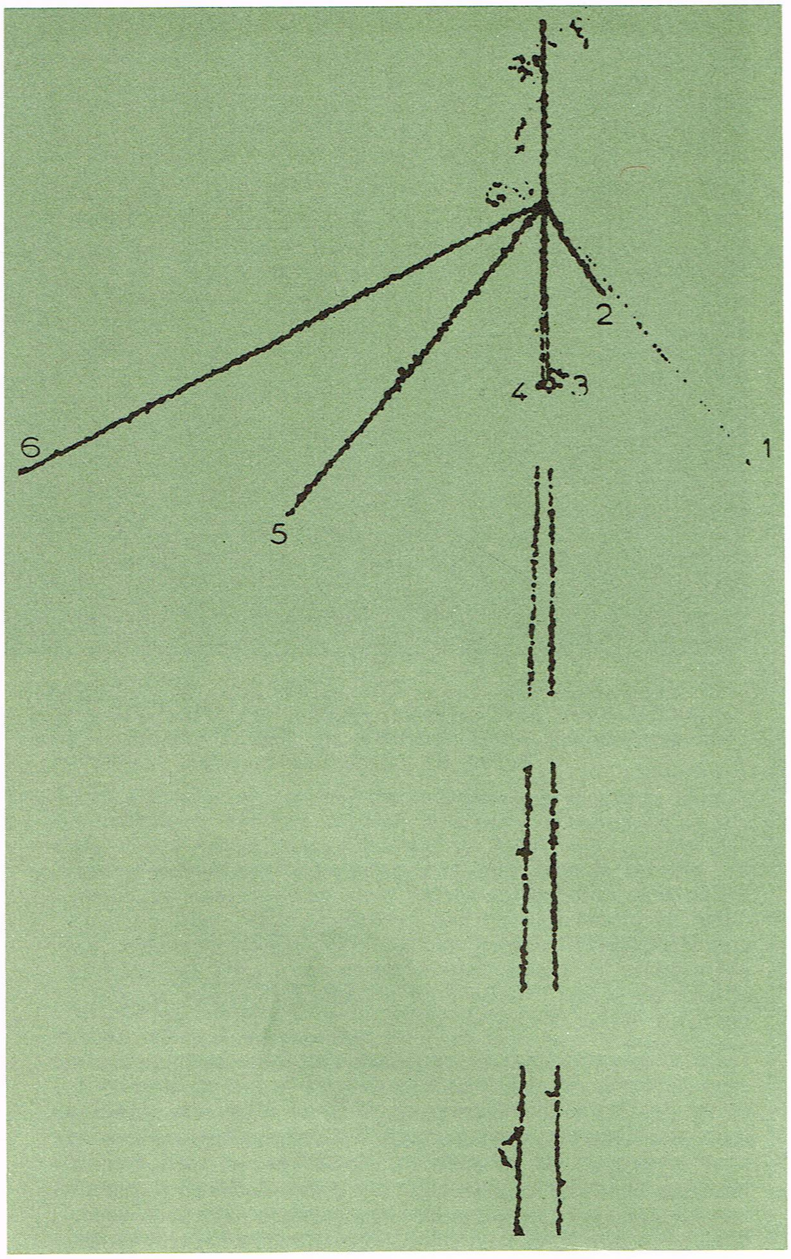
Phot. Laboratoire de physique, Ecole polytechnique.

par l'atmosphère, sont plus intenses, ils découvrirent ce nouveau méson qui interagissait avec la matière et dont le méson μ d'ailleurs n'était qu'un produit de désintégration.

Tout un domaine de la physique s'ouvrait avec le **méson** π et ses interactions avec les noyaux. Les recherches sur les rayons cosmiques devaient continuer de plus belle, amenant avec elles d'autres découvertes encore. Cependant, dès 1948, une année seulement après sa découverte, le méson π était créé au laboratoire!

C'est que, entre-temps, la technique des accélérateurs n'en était pas restée au stade du « Van de Graaff » et du cyclotron. En 1948, à Berkeley, aux Etats-Unis, des particules α étaient accélérées jusqu'à une énergie de 380 MeV et, envoyées sur une cible, produisaient des mésons π, détectés également par les émulsions photographiques. **L'époque des grands accélérateurs venait de s'ouvrir, reléguant au second plan les maigres apports des rayons cosmiques et leur substituant des faisceaux de particules d'énergies comparables.**

Les rayons cosmiques n'en devaient pas moins garder de l'intérêt pour eux-mêmes. Dans l'investigation du rayonnement cosmique primaire, avant sa dégradation dans l'atmosphère terrestre et à très haute altitude, on avait trouvé surtout des protons et d'autres noyaux d'éléments plus lourds existant aussi sur Terre. Mais d'où viennent ces particules : du Soleil, de notre galaxie ou d'ailleurs encore? Par quel mécanisme électrique ou magnétique sont-elles accélérées? Autant de questions qui n'ont pas encore reçu de réponse claire. L'ère des satellites devrait à ce point de vue élargir nos horizons et permettre une exploration systématique de l'espace, la découverte de la **ceinture de radiations de van Allen** qui entoure notre Terre à 1 000 km d'altitude étant peut-être à cet égard l'un des faits les plus marquants de ces dernières années.



Aspect des traces de particules dans une émulsion nucléaire.

les accélérateurs de haute énergie

Le « synchrotron à protons » du CERN, à Genève, est l'un des plus illustres représentants de ces puissants accélérateurs, puisque, actuellement, avec celui de Brookhaven, aux Etats-Unis, il fournit les protons les plus énergiques (30 000 MeV!). Les accélérateurs de haute énergie, sans être aussi importants, ne sont pratiquement plus des exemples isolés. Les types en sont nombreux et variés, mais ils peuvent se diviser en quatre grandes classes : **les accélérateurs à protons, linéaires ou circulaires; les accélérateurs à électrons, eux aussi linéaires ou circulaires.** Le principe général est le même pour tous, et c'est celui qui fut à l'origine du cyclotron : accélérer les particules par petits coups successifs, de façon à atteindre graduellement des énergies très élevées. Cette accélération est assurée dans le cyclotron par une tension électrique alternative de haute fréquence. Mais les possibilités de cette technique sont limitées, et la raison fondamentale relève de la théorie de la relativité qui trouverait ici, s'il en était encore besoin, une éclatante justification. Ainsi donc, dans le cyclotron, on oblige les particules à tourner en rond en les soumettant à un champ magnétique B; le rayon R du cercle qu'elles décrivent est alors relié à leur vitesse v ou, plus exactement, à leur impulsion $P = \frac{mv}{\sqrt{1 - \beta^2}}$, par la relation déjà vue $P = 300\ B\ R$.

Pour des vitesses v très faibles devant c, vitesse de la lumière, le terme relativiste $\sqrt{1 - \beta^2} = \sqrt{1 - \frac{v^2}{c^2}}$ se réduit pratiquement à 1, et l'impulsion P à mv; ainsi, avec l'augmentation de la vitesse, la particule décrit un cercle grandissant, mais met toujours le même temps pour faire un tour $\left(t = \frac{2\pi R}{v} = \frac{2\pi m}{300\ B}\right)$, et c'est ce temps qui impose la fréquence de la tension électrique d'accélération.

Que se passe-t-il pour des vitesses plus grandes? Le terme relativiste $\sqrt{1 - \beta^2}$ commence à intervenir notablement, la vitesse des particules s'approche de celle de la lumière et n'augmente plus que très lentement, beaucoup plus lentement que l'énergie et l'impulsion et donc que le rayon de la trajectoire. Ainsi les particules mettent plus de temps pour faire un tour et la tension accélératrice de fréquence fixe ne joue plus son rôle au bon moment. C'est la raison pour laquelle le cyclotron ne peut accélérer des protons que jusqu'à 10 MeV environ. Pour obtenir des énergies plus élevées, il a donc fallu modifier la technique primitive. Une première solution a été de faire varier la période de la tension accélératrice en la synchronisant avec le temps de révolution des particules : c'est cet appareil dérivé directement du cyclotron et appelé « synchrocyclotron » qui a créé à Berkeley les premiers mésons π.

Une deuxième solution a été de faire varier le champ magnétique B de façon à avoir toujours la même orbite circulaire. Cette solution s'impose aux hautes énergies. Les dimensions de la trajectoire circulaire et donc de l'appareil augmentent avec l'énergie : le diamètre de l'accélérateur du CERN est par exemple de 200 m et il est exclu de créer un champ magnétique sur la surface d'un cercle aussi grand.

Ce type d'accélérateur, appelé encore « synchrotron » (du grec *sun chronos*, « avec le temps »), se présente ainsi sous forme d'un grand anneau constitué d'une suite d'électro-aimants entrecoupés de temps en temps de blocs fournissant une tension accélératrice de haute fréquence. Les particules circulent dans un tube circulaire où des pompes entretiennent en permanence un vide de moins de 10^{-5} mm de mercure (1).

Un point important de la technique des accélérateurs a été passé jusqu'ici sous silence et il est d'ailleurs difficile de s'y étendre beaucoup. Pour que les particules reçoivent toutes en même temps le coup de fouet de l'accélération, il faut qu'elles restent groupées, et comme il existe toujours dans tout phénomène de petites perturbations, il faut faire en sorte que ces perturbations ne s'accroissent pas, mais qu'au contraire elles se détruisent d'elles-mêmes. C'est ainsi qu'un faisceau de particules beaucoup plus petit a pu être réalisé après la découverte du principe de « focalisation forte »; le tube où tournent les particules était de la taille d'un homme pour le synchrotron de 6 BeV de Berkeley; il a pu être réduit à des dimensions de 10 cm environ pour celui de 30 BeV du CERN.

(1) On se rappelle que la pression atmosphérique est de 760 mm de mercure.

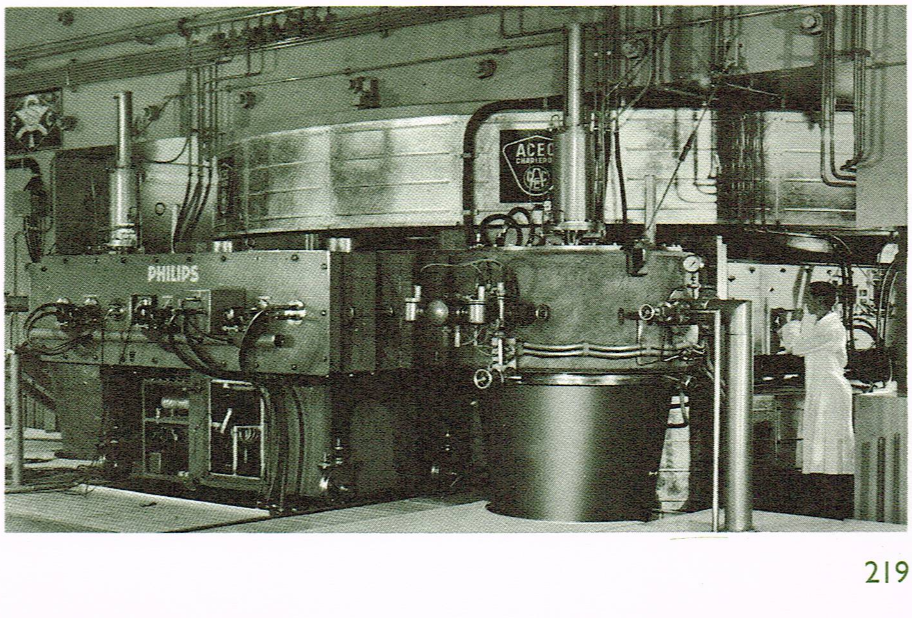
Après avoir fait des millions de tours en une ou deux secondes, les particules accélérées sont mûres pour l'exploitation. La technique la plus simple consiste alors à interposer une cible sur leur trajet. Les particules secondaires produites sont envoyées dans des détecteurs appropriés, après avoir été canalisées par des électro-aimants et des lentilles électromagnétiques en un faisceau dense et bien nourri. On commence maintenant, technique plus difficile, à utiliser directement le faisceau primaire lui-même.

Il existe beaucoup de types de détecteurs : compteurs électroniques de tout genre, émulsions nucléaires, chambres de Wilson. Mais celui qui, sans conteste, a donné les meilleurs résultats est un appareil qui a vu le jour après les accélérateurs eux-mêmes. Il s'agit de la **chambre à bulles,** dont le principe fut découvert en 1952 par le physicien américain D. Glaser.

Par certains côtés, la chambre à bulles s'apparente à la chambre de Wilson. Cette dernière contient un gaz sursaturé de vapeur d'eau. Ce gaz, qui est à une certaine pression, est soumis à une brusque détente qui provoque son refroidissement et la condensation de la vapeur d'eau. La chambre à bulles, elle, contient un liquide sous pression dont une détente brusque prot voque l'ébullition. Dans un cas comme dans l'autre, on peutrouver des conditions de température et de pression telles que la condensation de la vapeur ou l'ébullition du liquide se produisent sur le passage des particules électrisées. Ainsi les photographies révèlent la trajectoire desdites particules sous forme d'une suite soit de gouttelettes d'eau, soit de bulles de vapeur.

Mais on saisit tout de suite l'avantage des chambres à bulles : le liquide, milieu beaucoup plus dense que le gaz, se prête admirablement bien à la production d'interactions nucléaires. L'interaction elle-même est visible, dans sa simplicité ou sa complexité, tandis que dans la chambre de Wilson il faut recourir à l'artifice d'écrans métalliques et se contenter de voir les particules qui veulent bien en sortir. De plus, on sait faire fonctionner des chambres à hydrogène liquide, si bien que le mécanisme simple d'une interaction entre une particule et un proton (et non pas un ensemble complexe de nucléons que sont les autres noyaux) est directement à notre portée.

La chambre à bulles à hydrogène liquide est devenue l'instrument hors série de la physique des hautes énergies. On ne conçoit même plus un accélérateur qui n'en serait pas équipé. Une telle chambre peut d'ailleurs fonctionner aussi bien au deutérium qu'à l'hydrogène et servir à l'étude des réactions sur des neutrons comme sur des protons. Evidemment, les particules électrisées seules laissent la trace de leur passage



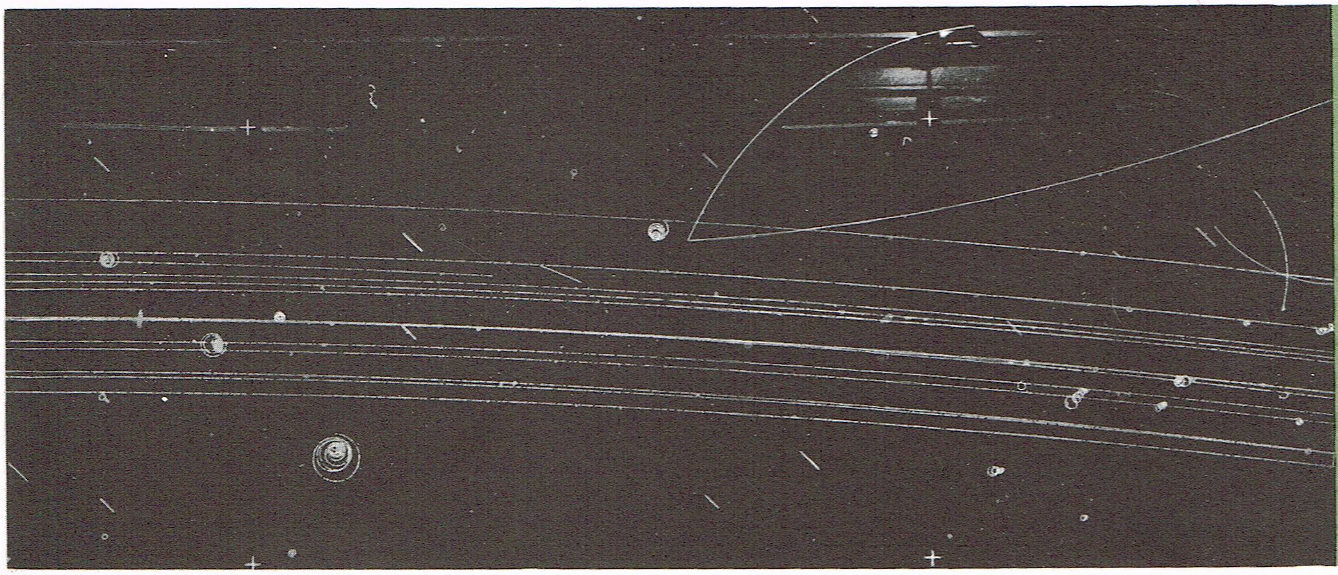
Synchrocyclotron du CERN. C'est ce type d'appareil qui a créé les premiers mésons π au laboratoire. Entourée par un électro-aimant, on voit la boîte dans laquelle les particules sont accélérées.

Phot. CERN.

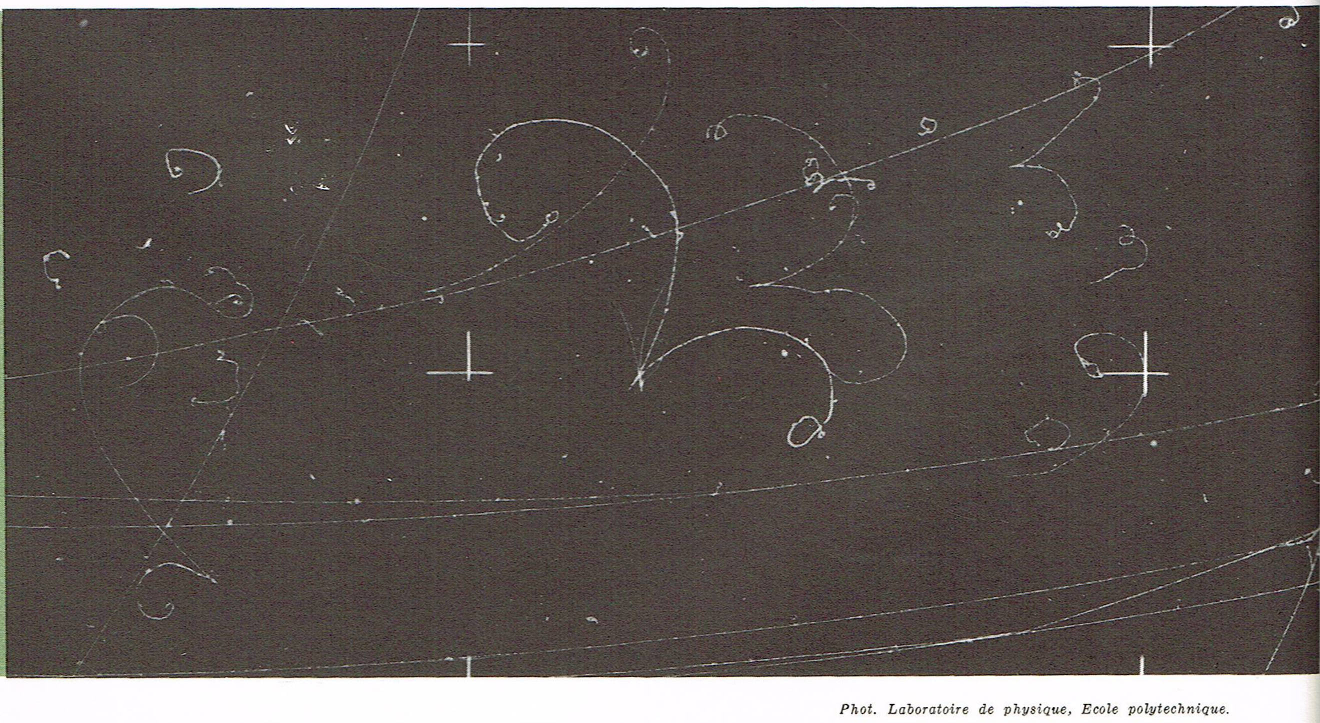
et les particules neutres — neutrons et γ en particulier — ne sont pas détectées. Aussi a-t-on construit, à côté des chambres à bulles à hydrogène, d'autres chambres, à liquide très dense, destinées avant tout à détecter la présence des γ de haute énergie : ceux-ci, à la faveur du milieu dense, se matérialisent facilement en paires d'électrons $e^+ + e^-$.

La chambre de Wilson pouvait fournir des photos intéressantes de rayons cosmiques à la cadence moyenne d'une ou deux par heure, à la condition toutefois d'être installée en haute montagne, là où les rayons cosmiques sont plus denses. La chambre à bulles fonctionne à la cadence même de l'accélérateur, toutes les deux secondes à peu près (la stabilité d'un gaz après détente étant plus difficile à obtenir que celle d'un liquide, la chambre de Wilson a un temps de répétition qui ne peut être inférieur à deux minutes), véritable machine à produire des photos, riches de contenu au surplus. On saisit la différence d'échelle et les problèmes que pose un tel rendement. Ce n'est pas seulement la marche de l'accélérateur et le bon fonctionnement de la chambre à bulles qui réclament une technique automatisée, mais c'est aussi l'exploitation des photos : **machines à dépouiller, machines à mesurer, machines à calculer sont autant d'outils indispensables. La physique des hautes énergies devient une entreprise de grande envergure, à l'image même des accélérateurs géants qui en sont la source; de plus en plus, elle requiert des spécialistes de toutes sortes et exige une étroite collaboration entre eux.**

De ce travail, commencé avec les rayons cosmiques, continué par les accélérateurs, quel est le résultat pour la compréhension des phénomènes nucléaires de haute énergie? Beaucoup de particules nouvelles, chacune avec son individualité propre, viennent enrichir un monde déjà complexe. Quelles sont les lois de cette étrange société, voilà qui n'est pas encore bien compris, mais les données expérimentales s'accumulent, et peu à peu l'exploration avance et un ensemble coordonné se dégage. Nous allons faire un tour d'horizon de ce domaine qui s'étend toujours et dont les limites se perdent dans l'avenir.

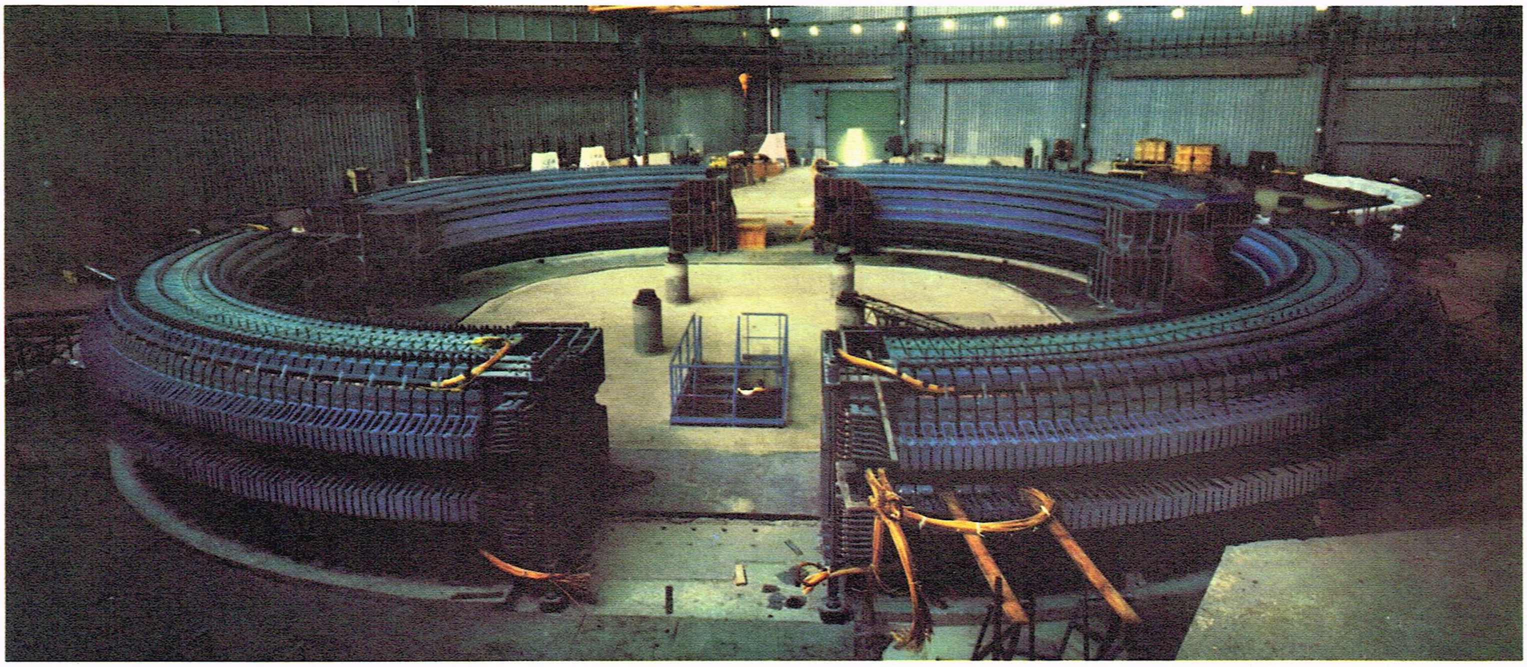


Faisceau de pions négatifs traversant une chambre à hydrogène. On voit, en particulier, l'une de ces particules disparaissant dans la chambre après interaction avec un proton du liquide.



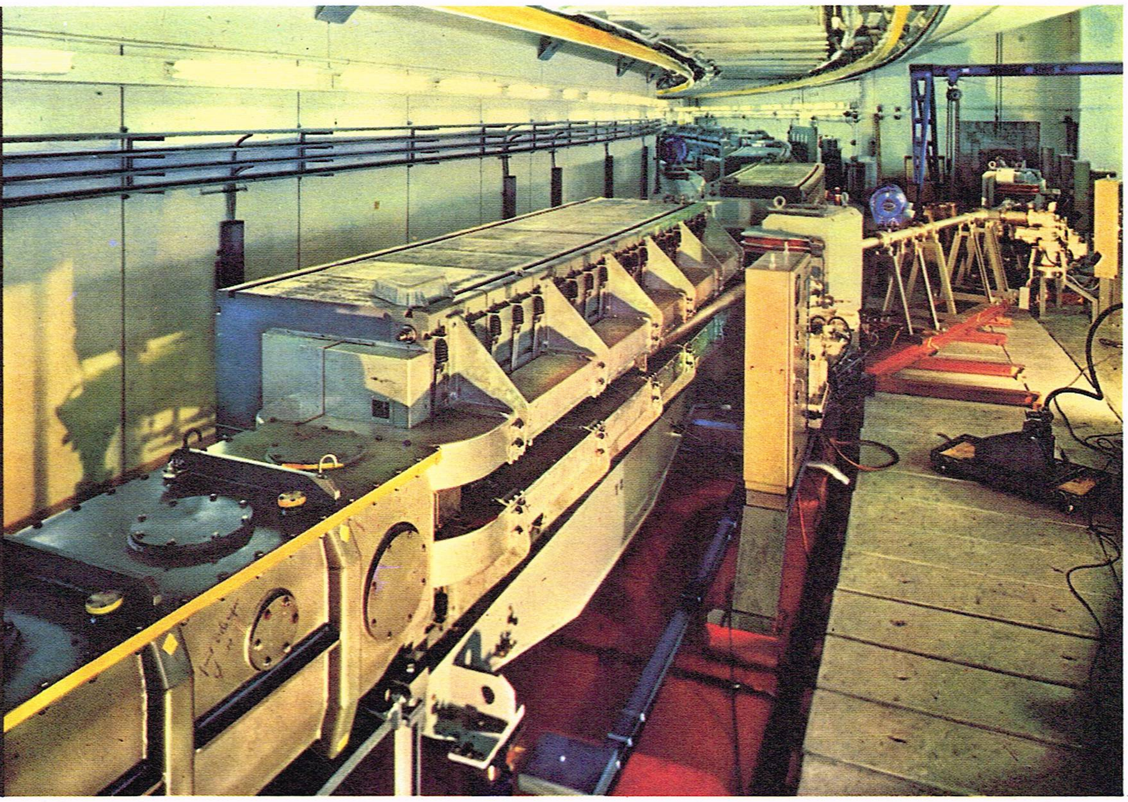
Faisceau de pions traversant une chambre à liquide lourd. La présence de rayons γ est révélée par leur matérialisation en paires d'électrons reconnaissables par leur spirale caractéristique.

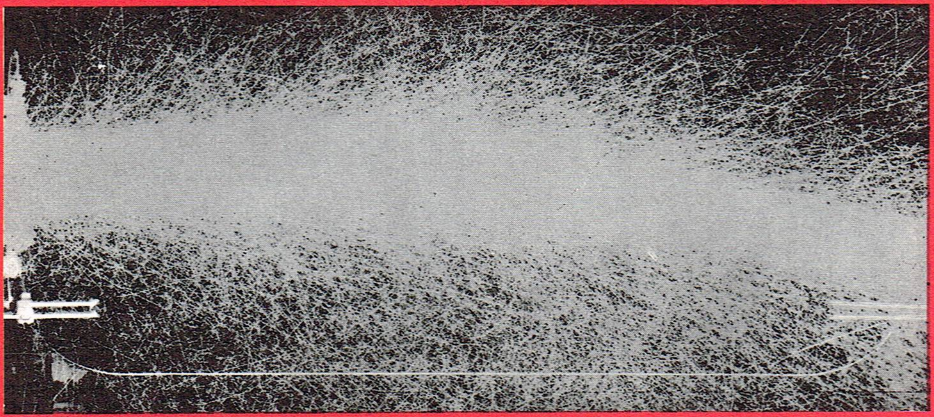
Phot. Laboratoire de physique, Ecole polytechnique.



Le synchrotron « Saturne », du Centre atomique de Saclay, permet de communiquer aux protons une énergie de 3 milliards d'électron-volts.
Doc. C.E.A.

Une section du synchrotron de 28 milliards d'électron-volts du CERN.
Doc. Brown-Bovery.





Un faisceau de particules accélérées par le synchrotron « Saturne » laisse les traces de son passage dans le liquide d'une chambre à bulles.
Doc. C.E.A.

le monde étrange des particules élémentaires

Les particules forment tout un monde : elles sont riches en nombre, riches en variétés, riches en caractéristiques. Il y a peu de temps encore, les physiciens croyaient que la liste des particules était close, mais de récentes découvertes ont fait rebondir la question. Il serait difficile de prétendre donner de ce vaste domaine une vue complète; bien des points importants seront même ici laissés de côté : espérons que l'ensemble y gagnera en clarté. Essayer de faire saisir la richesse de ce monde dans sa multiple variété, tel est notre but.
Les chapitres précédents, il est vrai, ont tous été consacrés à ces particules, mais le moment est venu de nous dégager des techniques expérimentales; prenons donc du recul, résumons nos connaissances. Leur synthèse nous fera déboucher sur les apports de la physique des hautes énergies.

le monde en trois particules ?

On peut dire que la matière est un assemblage complexe et varié de trois sortes de particules. Les protons et les neutrons forment le cœur de l'atome et les électrons (négatifs) mènent autour d'eux une course folle. Ces particules sont de très petits objets à l'image du noyau atomique lui-même dont le rayon est de l'ordre de 10^{-13} cm. Malgré cette petitesse, elles ont une structure interne très riche. Nous avons vu comment était née pour l'électron l'hypothèse d'un **spin** quantifié. Il faut se représenter non seulement l'électron, mais aussi le proton et le neutron, comme de petites boules tournant sur elles-mêmes, ce qui leur confère un moment cinétique propre appelé « spin ». Ce spin a d'ailleurs la même valeur $\frac{1}{2}$ pour ces trois particules, qui doivent ainsi présenter des similitudes de structure.

Pour expliquer les forces d'attraction ou de répulsion électriques, on a doté la nature de **charges électriques** sans évidemment savoir quelle réalité dernière se cache sous ce nom. Le neutron, comme son nom l'indique, est neutre électriquement et sa charge électrique est nulle. Proton et électron sont porteurs des charges respectives $+ e$ et $- e$ ($e = 1{,}6 \cdot 10^{-19}$ coulomb), qui s'annulent mutuellement. Ils ont toujours la même charge, appelée encore « charge élémentaire », car elle ne peut être fractionnée, pas plus que ne peut l'être un électron ou un proton (dans l'état de nos connaissances en tout cas!). En première approximation, une charge si petite peut être localisée en un point matériel, mais, la particule ayant des dimensions finies, il est logique de penser qu'elle occupe le volume même de celle-ci, comme si elle pouvait se diviser en une multitude de points matériels porteurs d'une charge plus élémentaire encore : on dit qu'il y a une répartition de la charge électrique à l'intérieur de la particule.

La particule tournant sur elle-même, si les charges se trouvent entraînées par ce mouvement, elles doivent donner naissance à de petits courants. Il en est bien ainsi : comme un courant, une particule est sensible à un champ magnétique, et cette influence se caractérise par un **moment magnétique,** qui peut être ainsi mesuré. C'est même la mise en évidence de ce moment magnétique qui a prouvé indubitablement l'existence d'un spin pour l'électron (expérience de Stern et Gerlach). On a mesuré également le moment magnétique du proton et du neutron, et ce dernier a des propriétés curieuses : sa charge est nulle, son moment magnétique ne l'est pas; il existe donc dans le neutron de petits courants internes, mais distribués de telle façon que la charge électrique totale reste nulle.

La distribution de charge à l'intérieur des nucléons s'explore systématiquement maintenant à l'aide d'électrons très énergiques, et c'est là l'une des principales utilisations des grands accélérateurs à électrons. Très énergiques, les électrons peuvent pénétrer au cœur même du nucléon, et la façon dont ils sont déviés renseigne sur la structure électrique interne du nucléon. Ils sont d'ailleurs seuls à pouvoir s'acquitter de cette besogne, puisque, au contraire des protons, ils ne manifestent pas d'interactions « nucléaires », qui, lorsqu'elles existent, sont beaucoup plus importantes que les interactions électromagnétiques et masquent complètement les effets de ces dernières.

Des trois caractéristiques citées ci-dessus (charge électrique, spin et moment magnétique), si la dernière n'est pas encore déterminée expérimentalement dans beaucoup de cas, sauf pour les particules n, p, e, μ, les deux premières, bien connues, donnent à toutes les particules des airs de famille; elles ont en effet des valeurs simples et très limitées, résumées ci-dessous :

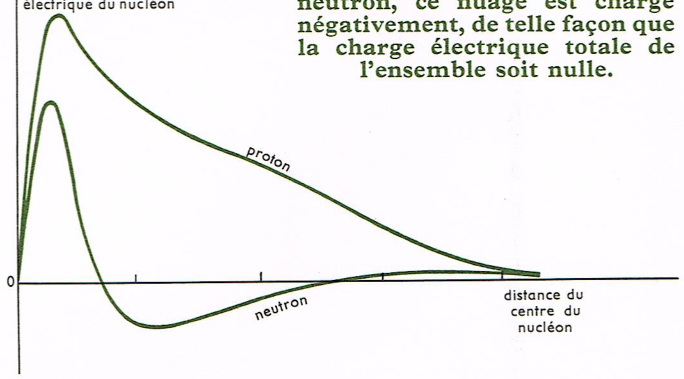
— charge électrique $+ e$ 0 $- e$;
— spin 0 $\frac{1}{2}$ 1.

Il est une autre caractéristique, essentielle elle aussi, mais que nous n'avons pas encore abordée. Nous savons bien que la matière est « pesante » et que c'est sa « masse » qui est responsable de son poids et de son inertie dans les mouvements. Eh bien, si proton, neutron et électron constituent la matière en dernier ressort, ils doivent se partager sa masse. Que cache exactement cette notion de masse? Nous en avons une idée plus précise depuis les travaux d'Einstein et ses principes fondamentaux, qui vont apparaître à propos du photon.

la masse d'une particule : du photon à la relativité

Si proton, neutron et électron suffisent à édifier un atome, il ne faut pas oublier que la réalité de l'univers englobe aussi la lumière et, plus généralement, les ondes électromagnétiques. A l'échelle atomique, ces dernières se présentent même comme de petits grains d'énergie, ou photons, et ressemblent à ces autres grains que sont nucléons et électrons. Une différence essentielle les en distingue pourtant. Quelle que soit l'énergie qu'ils transportent, qu'ils soient rayons γ ou rayons X, les photons ont toujours la même vitesse, celle de la lumière : $c = 300\,000$ km/s. Au contraire, les protons, par exemple, sont au repos dans un atome, mais il est possible de les accélérer et de leur donner avec une vitesse grandissante un surcroît

Densité de charge électrique du nucléon.



Les mesures les plus récentes portent à croire que, du point de vue électrique, le nucléon serait formé d'un cœur (partie gauche de la courbe) entouré d'un nuage plus ténu. Pour le neutron, ce nuage est chargé négativement, de telle façon que la charge électrique totale de l'ensemble soit nulle.

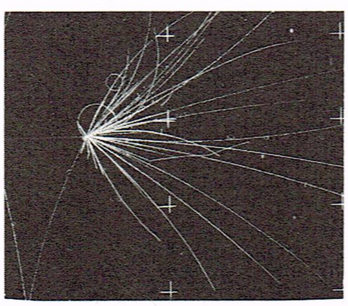
d'énergie qui va jusqu'à pouvoir briser des noyaux atomiques. **L'énergie, liée étroitement à la vitesse pour un proton, en est indépendante pour un photon.** Cet apparent paradoxe a reçu une explication claire grâce à la théorie de la relativité d'Einstein, qui vient ici en grande unificatrice.

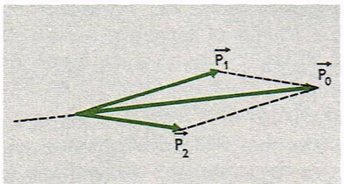
Pour les besoins de notre cause, cette théorie peut se résumer très brièvement : il n'est pas possible de dépasser la vitesse c de la lumière et une particule qui, au repos, a une masse m ne peut pratiquement pas atteindre cette vitesse, car son énergie deviendrait infinie. En d'autres termes et plus précisément, l'énergie totale d'une telle particule animée d'une vitesse v est donnée par la relation $E = \frac{mc^2}{\sqrt{1 - v^2/c^2}}$.

Au repos, la particule possède donc une énergie $E_0 = mc^2$, c'est-à-dire que sa masse m est le reflet de cette énergie intrinsèque. Si on lui communique une vitesse v, son énergie E augmente et deviendrait même infinie si la vitesse c de la lumière pouvait être atteinte (pour $v = c$, $E = \frac{mc^2}{0} = \infty$). Tel n'est pas, bien sûr, le cas du photon animé de cette vitesse c; sa masse doit donc être nulle et, à vitesse égale, son énergie peut alors varier ($E = \frac{0}{0}$ a une valeur mathématique indéterminée), devenant une caractéristique essentielle du type de photon et différenciant par exemple les rayons γ des ondes de radio. **Ainsi il est impossible d'imaginer un photon au repos ou avec une vitesse v inférieure à c, car l'énergie d'un tel photon serait nulle et, dans un univers où l'énergie est la réalité ultime, cela reviendrait à dire qu'il n'existerait plus.**

Avec cette théorie d'Einstein, le photon vient prendre sa place dans la grande famille des particules, qui se différencient par leur masse, c'est-à-dire par leur énergie interne. Certaines particules d'ailleurs, le proton et le neutron par exemple, ont des masses très voisines, presque égales, et, ne se différenciant que par leur charge électrique, peuvent être considérées comme les différents états d'une même particule. Si la liste des particules « élémentaires » s'en trouve ainsi raccourcie, il n'en reste pas moins vrai que la variété des masses existantes est grande et que la loi qui préside à leur répartition reste le grand problème à résoudre. Ni le pourquoi ni le comment de ces masses n'ont reçu encore le moindre éclaircissement.

Un méson π énergique a donné naissance à un grand nombre de particules. L'énergie se transforme en matière.





Représentation schématique de la conservation de l'impulsion $\vec{P}_0 = \vec{P}_1 + \vec{P}_2$.

création et disparition des particules

La grande idée qui n'a cessé d'inspirer l'étude du domaine de l'infiniment petit a été la recherche d'éléments inaltérables dans une nature où le changement est roi; mais l'inviolabilité des atomes et des noyaux atomiques est une notion qui a été détruite avec la découverte de la radio-activité. Elle n'existe pas plus au stade des particules. Un atome excité électriquement peut émettre des photons et il peut aussi les absorber : **un photon peut naître et il peut disparaître.** La radio-activité nous a montré qu'il pouvait sortir du noyau un électron qui n'y préexistait pas. La physique des hautes énergies nous montre fréquemment des interactions donnant naissance à des particules qui n'existaient pas au départ; beaucoup de ces particules connues sont même produites artificiellement et n'existent pas dans notre univers naturel de protons, neutrons, électrons et γ.

Ainsi l'objet de la physique des particules élémentaires ne se borne plus à l'étude de leurs propriétés internes, mais englobe aussi les lois qui président à leurs interactions. Ces dernières ne se font pas au hasard et il existe même des éléments de stabilité dont il faut essayer de connaître la raison.

Il faut d'abord insister sur le fait que toutes les interactions, qu'il y ait création ou non de particules, obéissent aux lois de la **dynamique relativiste,** dont celles de la dynamique ordinaire sont une approximation, valable aux vitesses faibles, et qu'en cela le monde des particules n'est pas un monde spécial. Un système de particules est caractérisé par son énergie totale E (énergie de masse + énergie cinétique) et par son vecteur impulsion totale $\vec{P}$. La dynamique relativiste nous dit que dans une interaction il doit y avoir conservation de l'énergie totale (E final = E initial) et conservation du vecteur impulsion totale ($\vec{P}$ final = $\vec{P}$ initial). Il y a une identité de lois entre la collision simple de deux particules et celle de deux boules de billard, et ce n'est pas par hasard que les trajectoires se ressemblent. Une collision avec production de nouvelles particules serait à comparer — image assez imparfaite, bien entendu — avec une collision de boules de billard dont une serait de verre et contiendrait des billes que le choc libérerait.

Ces lois de la dynamique relativiste valent d'ailleurs tout aussi bien pour des désintégrations radio-actives de noyaux que pour des désintégrations de particules, qui ne sont, somme toute, que des types particuliers d'interactions. Qu'il y ait collision de deux particules ou explosion d'une seule ne constitue qu'une différence superficielle; les mécanismes internes et les lois doivent être identiques. Il nous sera plus facile de préciser plusieurs d'entre elles en abordant la désintégration spontanée de certaines particules, ce qui viendra en même temps enrichir plus directement les connaissances que nous en avons.

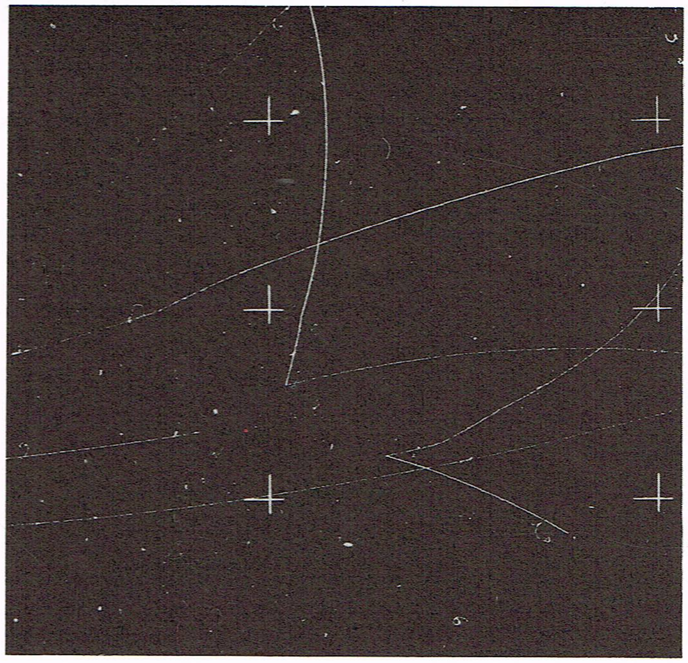
particules stables et instables

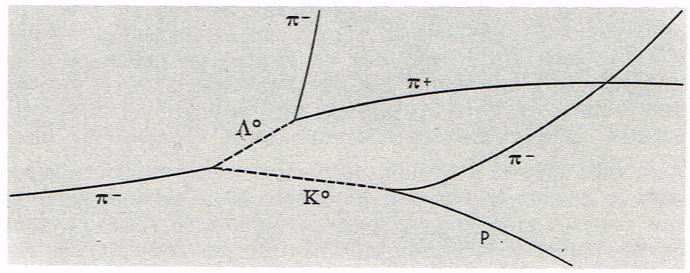
La radio-activité naturelle nous a fait connaître des noyaux qui se désintégraient spontanément, sans l'influence d'agents extérieurs. Ils se désintégreraient tout aussi bien dans le vide, et c'est bien en ce sens qu'il faut parler de stabilité ou d'instabilité des noyaux, de stabilité ou d'instabilité des particules. Il s'agit d'une propriété spécifique des particules, en dehors des liaisons qu'elles peuvent avoir avec le monde qui les entoure. Si un proton est isolé, il reste proton. Si un neutron est isolé, il ne reste pas neutron. Les particules connues ne sont pas stables pour la plupart, et le fait que quelques-unes le soient peut

Tableau des particules

Ce tableau indique uniquement les particules; toutes (sauf π et γ) ont des antiparticules symétriques. La masse est indiquée avec un signe ~ quand il y a de légères différences de masse (comme entre *n* et *p*) entre les particules de charges électriques de signe +, — ou 0. Quatre particules seulement sont stables.

	appellation		masse (en MeV)	spin	charge électrique	charge baryonique	charge leptonique	stabilité
	γ photon		0	1	0	0	0	oui
	ν neutrino		0	1/2	0	0	1	oui
	e électron		0,5	1/2	— 1	0	1	oui
	μ muon		105,6	1/2	— 1	0	1	
	π pion	~	140	0	+ 1, 0, — 1	0	0	
	K méson K	~	494	0	+ 1, 0	0	0	
N	*p* proton		938,2	1/2	+ 1	1	0	oui
N	*n* neutron		939,5	1/2	0	1	0	
hypérons (nucléon)	Λ^0		1115,4	1/2	0	1	0	
hypérons (nucléon)	Σ	~	1190	1/2	+ 1, 0, — 1	1	0	
hypérons (nucléon)	Ξ	~	1320	1/2	0, — 1	1	0	





Production de deux particules instables neutres Λ° et K° (appelées encore V° à cause de la forme que donne leur désintégration en particules chargées).

sembler si paradoxal qu'il est normal d'essayer alors d'expliquer cette stabilité. Quatre seulement sont d'ailleurs dans ce cas : **proton, électron, photon** et **neutrino** (pour l'instant nous nous limitons aux particules, mais les antiparticules ont des propriétés symétriques). Nous connaissons bien les trois premières de ces particules, mais c'est la première fois que nous rencontrons le neutrino, particule bizarre s'il en fut et qui passionne actuellement les physiciens. De charge électrique nulle, de masse nulle, elle ne provoque ni d'interactions nucléaires ni même d'interactions électromagnétiques, cela à la différence des photons. Sa découverte a été faite dans les phénomènes de la radio-activité β, de façon très indirecte d'ailleurs. En mesurant très soigneusement les énergies libérées dans les processus de radio-activité β, on s'aperçut d'un manque. **Si l'on voulait sauver le principe sacré de la conservation d'énergie, il fallait supposer l'émission, aux côtés de l'électron, d'une autre particule qui emportait l'énergie manquante.** Depuis lors, de l'existence de cette particule « ultra-neutre », des preuves directes ont été apportées, mais elles sont encore rares : un neutrino peut en effet très facilement traverser la Terre de part en part et en sortir inaltéré! Des expériences grandioses sont en cours par le moyen des accélérateurs du CERN et de Brookhaven pour obtenir de très hauts flux de neutrinos et pour pouvoir, à l'aide de tonnes de détecteurs, étudier leurs rarissimes interactions avec la matière.

Ainsi, dans toutes les radio-activités β, il y a, aux côtés d'un électron, émission d'un neutrino, et la désintégration type en est celle du neutron : $n \rightarrow p + \bar{e} + \bar{\nu}$ (1).

Evidemment, une telle désintégration n'est possible que parce que la masse du neutron est légèrement supérieure à celle du proton, et celui-ci ne peut pas pour cette raison se transformer en neutron. Mais pourquoi ne se désintègre-t-il pas en électrons ou en mésons?

Pour répondre à une telle question, il est intéressant de saisir la démarche de pensée du physicien : il invente des « nombres quantiques » nouveaux, dont il ne comprend pas nécessairement le sens, mais qui l'aident à trouver des règles de conservation et de stabilité. Pour cela, il procède souvent par analogie avec ce qu'il connaît déjà.

Une règle connue depuis longtemps et respectée par les interactions des particules est celle de la conservation des charges électriques : une particule de charge + e ne peut se désintégrer qu'en particules dont la somme des charges est + e. Cette règle de conservation vient expliquer ici pourquoi l'électron de charge — e, ou — 1 si l'on veut simplifier, ne se désintègre pas en ν ni en γ, de charge nulle.

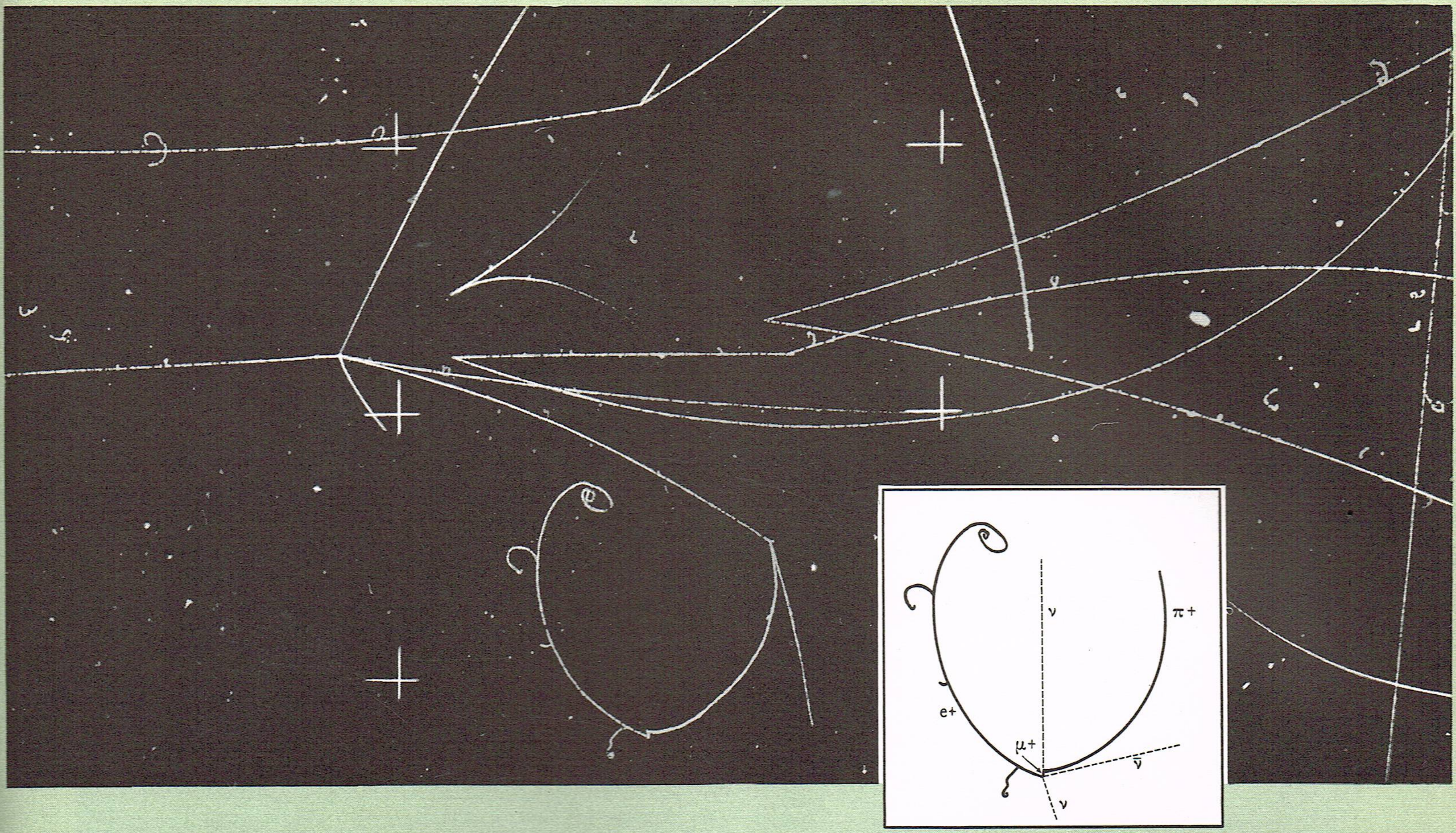
Pour expliquer la stabilité du proton, le physicien a inventé une nouvelle charge qui se conserve. Supposons en effet que le proton soit détenteur, à côté de sa charge électrique, d'une autre, d'un type spécial, et que les particules plus légères en soient dépourvues (pour cette raison, cette nouvelle charge a été appelée **baryonique,** du grec *barus*, « lourd »). Si celle-ci se conserve, le proton ne peut se désintégrer en particules plus légères; il est condamné à rester stable.

Dans cette distribution de charges, les particules légères ne sont d'ailleurs pas délaissées, car on a doté certaines d'entre elles — neutrino, électron et méson μ — d'une charge « leptonique » (du grec *leptos*, « léger »), de façon à expliquer entre autres l'émission conjointe d'un électron et d'un neutrino dans *tous* les processus de la radio-activité β.

La désintégration du neutron illustre par exemple cette conservation des charges leptoniques, aussi bien qu'électriques et baryoniques :

	n	$\rightarrow$	p +	e^- +	$\bar{\nu}$
charge électrique	0		1	−1	0
charge baryonique	1		1	0	0
charge leptonique	0		0	1	−1

(1) Nous verrons plus loin qu'il y a même deux neutrinos : un neutrino ν et un antineutrino $\bar{\nu}$.



Explication de la désintégration du pion positif (π^+) et du muon positif (μ^+).

Le π^+ s'arrête dans la chambre à hydrogène et se désintègre en μ^+ qui, s'arrêtant à son tour, se désintègre en e^+. La présence des coudes brusques est la marque de ces désintégrations en même temps que celle de l'émission simultanée de particules neutres. L'énergie cinétique du μ^+, mesurée par le parcours de cette particule, a toujours été trouvée identique, celle de l'électron, mesurée par la courbure de la trace, est, au contraire, variable : d'après les lois de la dynamique relativiste qui veulent que, dans une désintégration, énergies et impulsions soient conservées, ces faits indiquent clairement qu'une seule particule neutre doit être émise avec le μ^+, mais deux au moins avec l'électron. Comme des γ n'ont jamais été observés (ils auraient été détectés par leur matérialisation en paires d'électrons), ces particules neutres doivent être des neutrinos.

La règle de conservation des charges leptoniques vient expliquer la présence d'un seul, puis de deux neutrinos dans ces désintégrations successives. La charge leptonique du méson π^+ qui peut être créé dans des réactions comme $p + p \rightarrow p + n + \pi^+$ doit être nulle ; celles de l'électron et du méson μ, aux propriétés semblables, peuvent être supposées logiquement égales et les désintégrations doivent alors s'écrire :

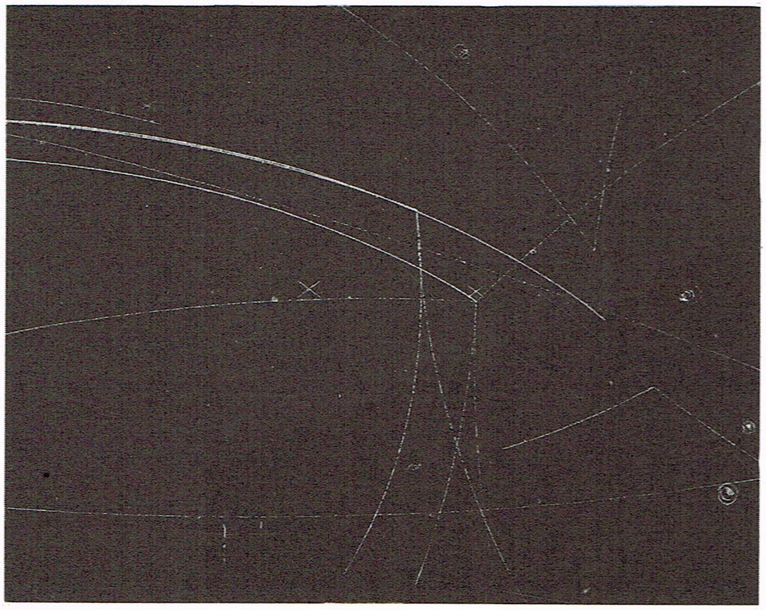
	$\pi^+ \rightarrow \mu^+ + \nu$	$\mu^+ \rightarrow e^+ + \nu + \bar{\nu}$
Charge leptonique	$0 \quad -1 \quad +1$	$-1 \quad -1 \quad +1 \quad -1$

Sur cet exemple, les affectations aux différentes particules de ces nouvelles charges peuvent paraître bien arbitraires, mais il existe beaucoup de processus de désintégrations et d'interactions (ci-dessus est expliqué le cas des mésons π et μ), et, par les recoupements qu'ils provoquent, on peut affirmer qu'il s'agit bien là, en plus des conventions de départ, d'une règle réelle de conservation. Que recouvrent exactement ces notions de charges baryonique et leptonique? Personne ne le sait vraiment, mais c'est un fait qu'elles viennent mettre de l'ordre dans les multiples jeux de ces particules et donner une explication simple à l'absence de certains processus. **Avant d'espérer accéder à une connaissance plus profonde des particules, il faut bien commencer par les classer et par faire œuvre de naturaliste.**

Les particules instables, qui se désintègrent en d'autres plus légères, stables ou instables, ont entre elles un trait commun qu'elles partagent avec les noyaux radio-actifs. Elles ne vieillissent pas, en ce sens qu'en un temps donné c'est toujours la même proportion d'entre elles qui se désintègre. Comme on avait défini pour les noyaux radio-actifs une période, on définit pour les particules une vie moyenne. Celle-ci est en général très courte, de l'ordre de 10^{-10} s (pour le neutron seul, elle atteint le temps exceptionnellement long de 14 minutes). Sa mesure permet de connaître l'intensité des mécanismes qui ramènent finalement tout ce monde vers la stabilité des quelques particules qui forment notre univers habituel.

les antiparticules

Nous avons déjà signalé l'existence de deux sortes de neutrinos, ν et $\bar{\nu}$, et le moment est venu de parler des antiparticules. Leur existence fut avancée la première fois par Dirac en 1929, à propos de l'électron positif, qui fut découvert un peu plus tard. Cet électron a exactement même masse que l'électron ordinaire, mais sa charge est positive. En face de chaque



Annihilation de protons par des antiprotons. Les antiprotons entrent dans la chambre à bulles (traces épaisses) et s'y arrêtent en capturant un proton du liquide. Une première annihilation donne deux V°. Une seconde donne quatre particules chargées, soit deux π^+ et deux π^-, et la réaction s'écrit :

$$p + \bar{p} \rightarrow 2\pi^+ + 2\pi^- + \pi^\circ.$$

Dans certains cas, cette même réaction se passe d'abord par l'intermédiaire de deux particules à vie extrêmement courte, suivant le schéma :

$$p + \bar{p} \rightarrow \omega^\circ + \rho^\circ$$
$$\rho^\circ \hookrightarrow \pi^+ + \pi^-$$
$$\omega^\circ \hookrightarrow \pi^+ + \pi^- + \pi^\circ$$

particule on connaît actuellement une antiparticule, c'est-à-dire une particule de même masse et de même spin, mais dont toutes les autres caractéristiques, charges électrique, leptonique, baryonique, moment magnétique, sont inversées. Seules des particules comme le photon ont une symétrie telle qu'elles se confondent avec leur antiparticule hypothétique. Ainsi on connaît l'antiproton et l'antineutron, antiparticules qui correspondent au proton et au neutron, et même le neutrino ν et l'antineutrino $\bar{\nu}$, dont la différence de charge leptonique, + 1 et — 1, vient très bien expliquer, par exemple, la désintégration des mésons π et μ.

Une certaine science-fiction a beaucoup parlé d'une bombe à antiparticules. La raison en est la suivante : si des antiprotons (de charge baryonique — 1) rencontrent des protons (de charge baryonique + 1), ces particules ont de grandes chances de se détruire mutuellement et de se transformer par exemple en deux mésons π (de charge baryonique 0). L'énergie libérée par différence de masse ([2 × 938] — [2 × 140] MeV $\simeq$ 1 600 MeV) est énorme, même comparée aux 200 MeV libérés par la fission d'un noyau d'uranium. Là il s'agissait seulement de transformations de noyaux; ici c'est la destruction même de la matière par l'antimatière qui est en jeu. Heureusement, notre monde n'est fait que de matière et sa stabilité n'est pas compromise. S'il a fallu la construction à Berkeley du synchrotron à protons de 6 BeV pour créer les premiers antiprotons et antineutrons (il fut d'ailleurs construit pour cela), c'est que, aux côtés d'un antinucléon, un autre nucléon doit nécessairement être créé (règle de la conservation des charges baryoniques), et l'énergie dépensée est sans commune mesure avec celle libérée par annihilation des quelques antinucléons produits. On ne voit vraiment pas comment une bombe à antiparticules pourrait sortir du domaine de la science-fiction.

En revanche, il n'est pas interdit de rêver à d'autres mondes formés d'antimatière et d'imaginer ce qui arriverait en cas de collision avec la Terre...

En tout cas, il est curieux de constater que **peuvent exister deux mondes symétriques, mais incompatibles entre eux, et qu'un monde d'antimatière serait aussi stable qu'un monde de matière.**

des particules qui n'en sont pas

La liste des particules comprend celles qui, stables ou instables, sont acceptées comme telles aujourd'hui. Même instables, elles n'en sont pas moins d'ailleurs « visibles » en général : elles laissent une trace dans une chambre à bulles; elles parcourent une certaine distance avant de se désintégrer.

Les propriétés de ces particules ont retenu et retiennent encore l'attention des chercheurs, qui en avaient trouvé certaines « étranges » parce qu'elles ne se comportaient pas comme ils l'eussent imaginé. Cependant, si la création de ces particules rend souvent simples les interactions à énergie relativement peu élevée, il en va autrement à de plus grandes énergies comme celles qui sont fournies par les accélérateurs du CERN ou de Brookhaven, ou même par celui de Berkeley. Des interactions, il sort souvent des particules à foison; on a alors essayé de les grouper deux à deux, trois à trois pour voir si elles n'avaient pas d'affinités spéciales entre elles. C'est ainsi qu'en faisant leur bilan d'énergie et d'impulsion, on a trouvé que tout se passait parfois comme si des particules transitoires, à vie extrêmement courte, avaient d'abord été produites.

Bien plus, si la masse de la supposée particule est calculée, ce n'est pas une valeur précise qui est trouvée, mais une répartition autour d'une valeur moyenne. La particule — si particule il y a — est si instable, que sa masse, c'est-à-dire son énergie interne, n'est même plus bien définie.

La largeur de la courbe obtenue, appelée encore courbe de résonance, par analogie avec les courbes de résonance bien connues dans les phénomènes physiques à notre échelle, constitue une indication de la vie moyenne de la particule nouvelle. Il se trouve que celle-ci est souvent si courte qu'en pratique la nouvelle particule, créée dans un volume de rayon de 10^{-13} cm qui est celui des particules interagissantes, n'en sort même pas « vivante ». Peut-on alors parler encore de particule? L'esprit répugne à appeler ainsi un objet qui disparaît presque avant d'être créé, et on préfère parler de « résonances ». Peut-être d'ailleurs s'agit-il simplement d'une propriété de l'interaction elle-même, puisque nous sommes en plein dans son domaine? Peut-être aussi cette découverte des résonances va-t-elle faire évoluer nos idées sur la définition même d'une particule, sur sa structure interne, sur les mécanismes qui jouent et qui la rendent plus ou moins stable? Peut-être nous donnera-t-elle enfin la clé de la nature profonde de la masse?

Comme on le voit, la physique des particules élémentaires est tout entière tournée vers l'avenir, et l'avenir devrait apporter la réponse à de nombreuses questions qui se posent sur ce monde bien mal connu encore, mais riche de promesses.

UNE JEUNE PHYSIQUE
probabiliste

La physique « classique » qui a régné tout au long du XIXe siècle et y a obtenu les succès les plus éclatants contemplait un univers particulièrement serein : espace bien uniforme, digne d'un « jardin à la française », indifférent à son contenu; temps impassible, s'écoulant régulièrement et indéfiniment; possibilité — surtout — d'une connaissance instantanée et illimitée. Olaüs Rœmer avait bien montré, dès le début du XVIIIe siècle, que la vitesse de la lumière était finie; mais il a fallu attendre deux siècles pour qu'Einstein tirât les conséquences d'une vérité longtemps inaccessible : si nul message ne pouvait se transmettre instantanément, notre définition de l'espace et du temps devait changer. Et ce fut, en 1905, la théorie de la relativité. Réconciliant mécanique et optique, elle constituait la vraie synthèse et le couronnement de la physique classique.

Mais de plus rudes assauts se préparaient qui n'allaient pas tarder à ruiner une conception si olympienne du monde. Trois ordres de réflexion, d'apparence bien différents, convergeaient finalement pour proposer de l'univers une vision probabiliste, fort peu intuitive, fort ardue mathématiquement, mais beaucoup plus logique,... et admirablement vérifiée par l'expérience : ce sera la physique probabiliste moderne, dont les deux plus beaux fleurons sont la théorie de l'information et la mécanique quantique (ou mécanique ondulatoire), permettant au physicien de reconstruire le monde de son expérience et de déterminer l'architecture de la matière, et celle du Cosmos même, à partir des quelques particules et des quelques formules qu'il a si difficilement synthétisées.

Aucune connaissance ne peut être infiniment précise. A cela, trois raisons :

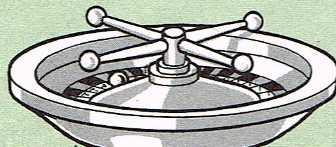
Toute observation ou toute expérience perturbe l'objet ou le champ que l'on désire connaître : on ne peut s'informer sans déformer, et il faut renoncer à préciser exactement ce que l'on a mesuré. Et aussi à refaire une expérience qui n'est jamais reproductible;

Tout système physique ne peut prendre qu'un nombre fini d'états : la matière ne connaît pas le continu et n'obéit qu'à des lois de quantification. L'existence d'un seuil de sensibilité, d'un seuil de saturation et d'un seuil différentiel interdit au résultat d'une mesure d'être quelconque : il ne peut prendre qu'un nombre fini de valeurs. Plus infortuné encore que son collègue mathématicien, qui peut, à défaut d'une connaissance infiniment précise de π (il lui faudrait être immortel pour avoir le temps nécessaire de calculer une infinité de décimales!), en avoir une connaissance aussi précise qu'il le désire (100 000 décimales actuellement, plusieurs milliards s'il veut y mettre le prix!), le physicien rencontre ici un seuil qu'il ne peut aucunement franchir, car il marque la limite de ce qui a un sens physique et de ce qui n'en a plus aucun;

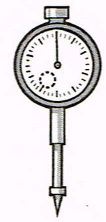
Il ne suffit pas de mesurer, il faut encore que le résultat nous soit transmis par un canal physique quelconque — y compris un organe des sens — et que nous puissions le comparer à d'autres résultats, le situer dans un ensemble et le conserver dans une mémoire (cellule cervicale), nécessairement limitée elle aussi. Et la communiquer en temps utile à toute personne intéressée, puisque la science ne vit que d'échanges. Or, le chapitre suivant nous apprendra précisément que tout canal physique ne peut écouler qu'une quantité finie d'informations, en un temps fini; et encore est-ce au prix d'une nécessaire dégradation. Il faut définitivement renoncer à toute connaissance infiniment précise. Et vivre dans un monde probabiliste, avec le seul espoir de cumuler le plus efficacement possible le plus d'informations possible.

Ce sont ces différentes remarques que nous allons développer et approfondir, en un exposé que nous avons voulu rigoureux, mais de présentation suffisamment vivante pour que le lecteur ressente que c'est par une réflexion critique sur les faits les plus banals de la vie que s'élabore la science — et non en niant cette expérience quotidienne, ni en essayant de torturer telle observation isolée ou telle équation. Curiosité, patience, esprit critique et bon sens sont bien encore les qualités les plus essentielles du scientifique.

Puissent aussi ces quelques pages donner les clés du monde moderne et reposer un peu le lecteur avant deux chapitres beaucoup plus austères par leurs mathématiques.



THÉORIE DES PROBABILITÉS



la réponse à une excitation n'est jamais susceptible que d'un nombre fini (quoique généralement très grand) d'états

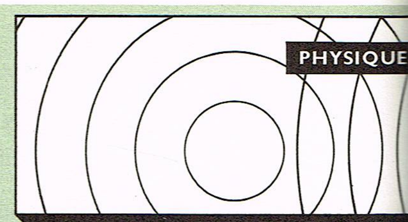
passage du discret au continu

N=3

N=10

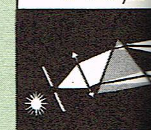
N=30

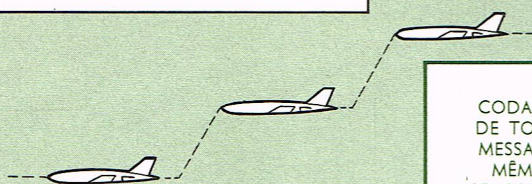
N=∞



électro-magnétisme

décomposition et analyse





CODAGE DE TOUT MESSAGE MÊME SENSORIEL

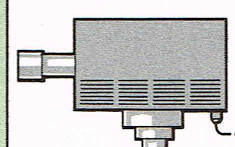
AGITATION THERMIQUE BRUIT DE FOND

ANALYSE D… THÉ… MATHÉM… DES FON… SUPPORTS

L'INFORMATION est toujours limitée, il faut l'utiliser efficacement

blablablablabla

tout mes… transmet… fini et au … dégra…



statistiques et théorie de l'estimation

recherche opérationnelle (science du gouvernement)

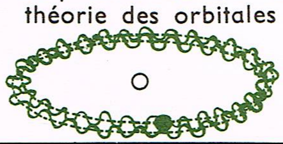


Louis de Broglie

MÉCANIQUE QUA… (ONDULATO…

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\psi}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2\psi}{\partial\varphi^2}-\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2}$$

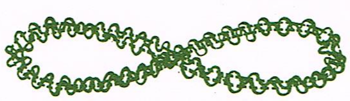
ondes de probabilités de l'électron théorie des orbitales



limitation de la connaissance par les formules d'Heisenberg

$\tau\Delta E\sim h \quad \Delta x\Delta p_x\sim h \quad \Delta y\Delta p_y\sim h \quad \Delta z\Delta p_z\sim h$

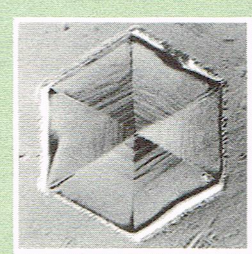
liaisons chimiques, état des électrons (valence, covalence) - énergie d'échange



ARCHITECTURE DE L'UNIVERS (reconstruction du monde à partir de particules élémentaires)

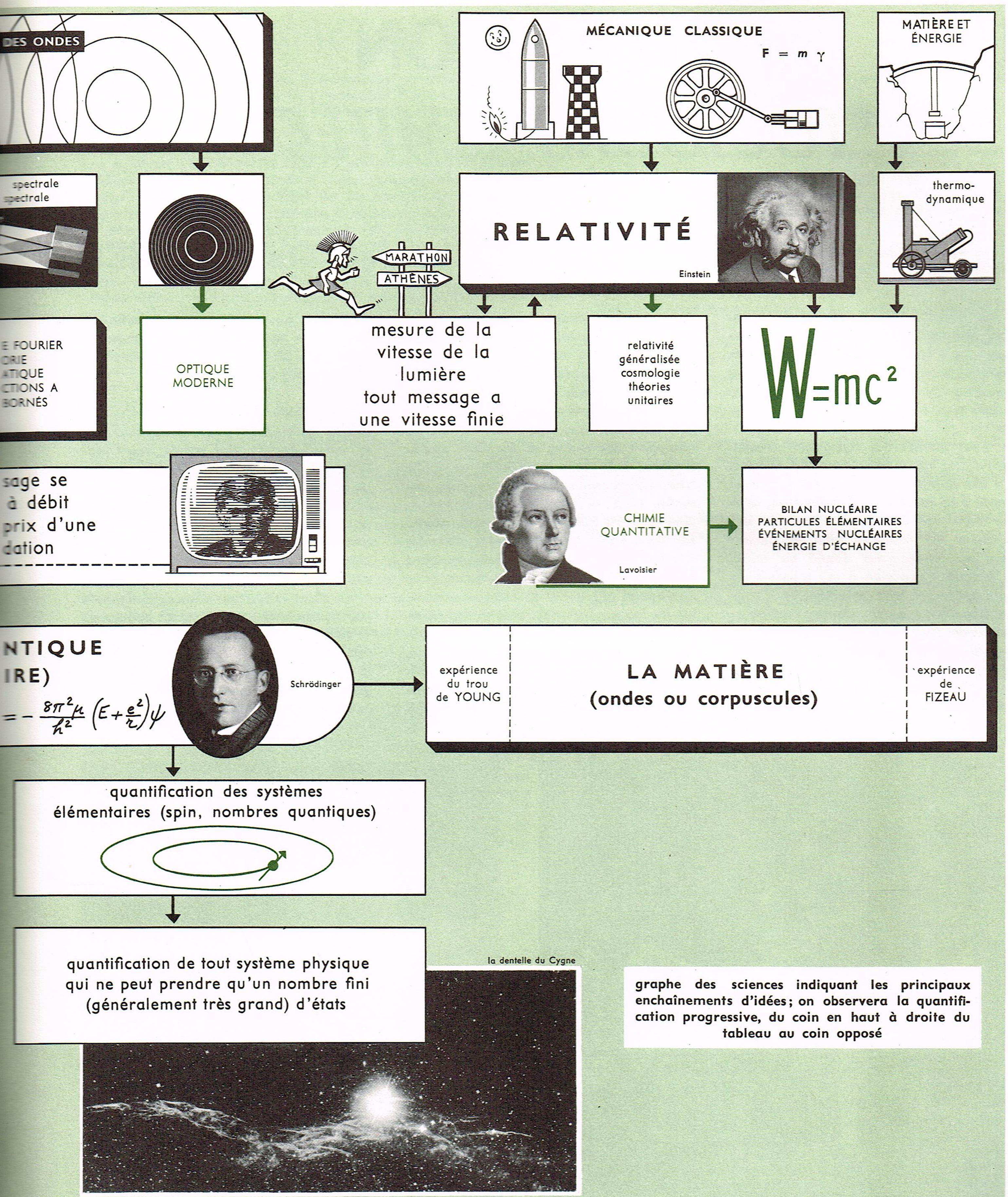


micromodule



microcristal

Phot. Keystone, E. Marton, Larousse, Fairchild, In…



graphe des sciences indiquant les principaux enchaînements d'idées; on observera la quantification progressive, du coin en haut à droite du tableau au coin opposé

UNE INFORMATION *TOUJOURS LIMITÉE*

par

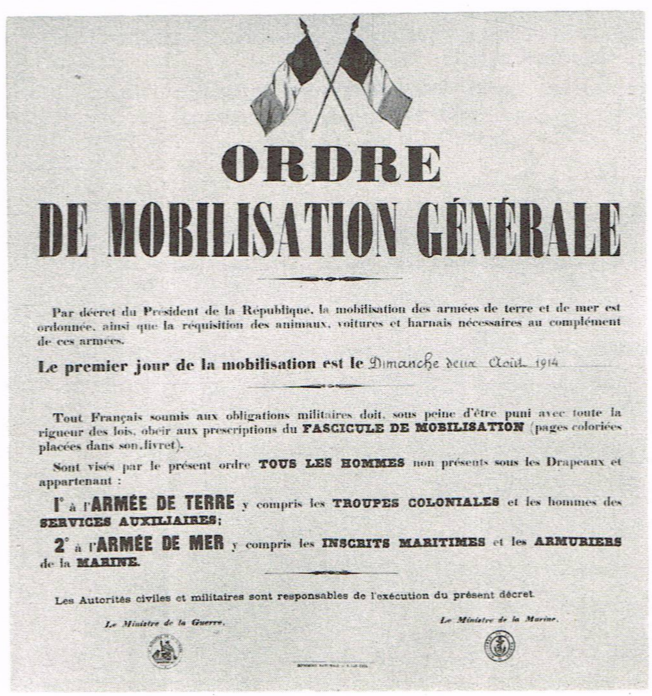
Jacques CLAVIER

« Par décision du ministre de l'Éducation nationale, la durée des grandes vacances a été arrêtée comme suit : les cours vaqueront du samedi 29 juin au soir jusqu'au dimanche 15 septembre inclus. »

Ce petit libellé, qui comprend trente-trois mots, est paru au Journal officiel. Il a été publié dans les journaux, commenté à la radio et à la télévision, affiché dans les écoles. La fixation des vacances est toujours bien accueillie; c'est un rendez-vous donné au soleil. La première réaction à ces trente-trois mots est aussi tout affective et d'imagination. Sans aucun doute, l'administration des Postes et Télécommunications va-t-elle constater dans les jours qui suivent une certaine augmentation de son trafic postal à destination des plages ou de la montagne. Mais, le 30 juin, aussi bien, hélas! que le 15 septembre, un ensemble de faits étonnants, tout à fait physiques et très mesurables, s'exprimera sur tout le territoire. Des millions d'ampoules électriques s'éteindront plus tard la veille, et des millions de réveils sonneront plus tôt. Puis, dans la matinée, une quantité d'essence comprise entre cinq et dix mille tonnes brûlera, non sans parcourir un cycle thermodynamique, réglé — plus ou moins bien — par nos constructeurs. Une certaine quantité de gomme restera sur les routes, principalement dans les virages, ainsi qu'un certain nombre (que l'on peut d'ailleurs prévoir avec une bonne précision) de voitures avec la vie de leurs conducteurs et de quelques passagers. La perturbation ne se bornera pas au trafic routier ou ferroviaire. Le mouvement de population entraînera le mouvement corrélatif des marchandises; le flux des denrées périssables sera détourné de sa destination traditionnelle. La localisation nouvelle des demandes de biens de consommation aura une incidence très profonde sur l'économie générale.

Il n'est d'ailleurs point besoin d'une si longue littérature pour déclencher des processus encore plus vastes, voire plus sinistres. L'affichette ci-dessous en est un bon exemple, et peu recommandable. Le titre, à lui seul, et la date eussent suffi pour provoquer le déplacement coordonné de millions d'hommes, dont la chance de survie fut a posteriori un peu supérieure aux deux tiers. Elle fut à l'origine d'une dépense d'énergie évaluée principalement en millions de tonnes de T. N. T. L'amplification, il est vrai, a fait de nos jours de grands progrès, et la pression d'une manette libère « en gros » la même quantité d'énergie avec un délai allant jusqu'à une demi-heure, selon l'objectif.





ORDRE
DE MOBILISATION GÉNÉRALE

Par décret du Président de la République, la mobilisation des armées de terre et de mer est ordonnée, ainsi que la réquisition des animaux, voitures et harnais nécessaires au complément de ces armées.

Le premier jour de la mobilisation est le Dimanche deux Août 1914

Tout Français soumis aux obligations militaires doit, sous peine d'être puni avec toute la rigueur des lois, obéir aux prescriptions du FASCICULE DE MOBILISATION (pages coloriées placées dans son livret).

Sont visés par le présent ordre TOUS LES HOMMES non présents sous les Drapeaux et appartenant :

1° à l'ARMÉE DE TERRE y compris les TROUPES COLONIALES et les hommes des SERVICES AUXILIAIRES;

2° à l'ARMÉE DE MER y compris les INSCRITS MARITIMES et les ARMURIERS de la MARINE.

Les Autorités civiles et militaires sont responsables de l'exécution du présent décret

Le Ministre de la Guerre. Le Ministre de la Marine.

Phot. Doisneau-Rapho, Larousse.

qu'est-ce donc que l'information ?

On voit quelles conséquences, physiquement mesurables, peut avoir la diffusion d'un ordre ou l'apposition d'une affiche. Bien entendu, l'étendue du désastre n'est pas proportionnelle à la surface de l'affiche ni à la force avec laquelle l'ordre a été proféré. Il convient de réfléchir à cette constatation, quelque peu comique : **les travaux mis en jeu dans les explosions nucléaires ou classiques sont bien proportionnels aux quantités de matières fissiles ou de poudre (comme les kilowatts-heures d'une centrale à la hauteur de chute et au débit), mais non à la grosseur des boutons de commande.** La typographie de l'affiche comme le ton de la déclaration n'ont que peu d'importance (dans certaines limites toutefois : une écriture anormale comme une élocution hors mesure amèneraient un élément supplémentaire à l'information apportée au public).

Les exemples précédents ont un caractère grandiose, général, mais exceptionnel. On peut montrer, en revanche, combien une foule d'avertissements de types les plus divers peuvent régler, et quelquefois dans les moindres détails, le cours de notre vie quotidienne.

La journée commence pour nous avec la sonnerie de notre réveil; l'écoute du bulletin météorologique qui précède les informations peut modifier (surtout chez les auditeurs naïfs) le choix de leur vêtement ou l'itinéraire de leur week-end. La réflexion sur la situation politique s'ajoute à celle des jours précédents. Elle se traduira par un comportement déterminé lors de situations spécifiques comme une consultation électorale. L'énoncé des faits divers amuse, indiffère ou révolte.

Vient ensuite la lecture du courrier. Beaucoup de lettres n'expriment que la banalité. D'autres, au contraire, sont très explicites : une feuille de papier, quelques lignes tracées, et quels que soient l'encre, la plume, le style ou l'écriture, le poids de la lettre, la couleur du papier, vous vous rendez à une heure précise en un lieu précis, ou passez le temps à préparer la réception d'un correspondant. Ce dernier, alors, utilise volontiers un télégramme.

ARRIVERAI JEUDI 8 GARE MONTPARNASSE A 17 HEURES AMITIÉS JEAN

Ce message, au libellé concis, a rigoureusement les mêmes conséquences qu'une lettre annonçant que Jean prend le train pour arriver à telle heure à la gare du Montparnasse, et qu'il fait part de ses amitiés. En fait, s'il eût payé son message au nombre de lettres ou à la ligne, et non aux mots, Jean aurait probablement réussi à s'exprimer en peu de signes :

ARR JEUD 8 MTPARN 17 H AMIT JEAN

ou encore moins selon son avarice ou son astuce.

Sur cette contraction repose le principe même des petites annonces, où, pour des raisons économiques, l'objectif est de signifier le plus avec le minimum de mots : elles sont « peu redondantes ».

Mais la journée va bientôt débuter, et généralement par la lecture du journal. Les titres attirent d'abord (à moins que ce ne soit l'article du chroniqueur préféré).

Le lecteur se rend compte, toutefois, que certaines photographies (accidents, réunions, etc.) l'informent mieux que le texte qui les accompagne. Telle caricature réussie dépeint excellemment une situation ou les traits marquants d'un homme d'Etat. Peut-être le dessin est-il à l'écriture ce que les télégrammes sont aux lettres. Le tableau des résultats du tirage de la Loterie nationale est remarquable par sa concision. Un chiffre changé, et le numéro gagnant devient perdant. Notons que les réactions à cette suite assez courte de chiffres sont très diverses : indifférence de ceux qui n'ont pas acheté de billet, déception, d'ailleurs chronique, de la foule des perdants, exaltation de l'éventuel gagnant.

On perdrait son temps à énumérer les occasions où l'homme, dans sa journée de travail, ne cesse d'être « informé » : le médecin écoute son patient et l'exposé de son cas, s'informe par l'auscultation, lit les radiographies, interprète les électrocardiogrammes. Le fraiseur ou le tourneur examinent le dessin d'une pièce. Les vitesses du tour ou de la fraiseuse, la dépense d'énergie, l'échauffement des pièces, le volume des copeaux en découlent naturellement.

Cependant, par téléphone ou note de service, le directeur technique règle une mise en fabrication, déterminant l'activité de centaines, voire de milliers d'hommes; le technicien modifie par télécommande la trajectoire d'un engin spatial, le syndicaliste harangue ses sympathisants, dont l'action consécutive peut avoir d'importants effets sur tel secteur de l'économie. Il semble bien que ces rapports, lettres, ordres, discours, appels téléphoniques soient la manifestation la plus directe, pour l'individu, du monde extérieur.

Ainsi disons-nous couramment que nous avons été « informés » par la radio, la télévision, les journaux, les orateurs, par nos interlocuteurs, par la lecture des livres, des lettres, des messages.

Nous entendons par là que nous ne nous comportons plus après comme nous l'eussions fait avant. **L'information a pour première conséquence une motivation meilleure de nos gestes et de nos pensées.** Car, même au cas où la lecture d'un cours de bourse ne provoque chez le spéculateur aucune dépense mesurable d'énergie (il a décidé de ne pas donner d'ordres), elle ne l'a pas laissé indifférent et a conditionné son humeur.

Le même journal qui propose les dernières nouvelles n'a que peu d'intérêt quelques jours plus tard. Les mêmes titres, les mêmes phrases, les mêmes chiffres et lettres ne provoquent plus que des réactions de souvenir, avant de se perdre dans l'oubli ou servir beaucoup plus tard de documents historiques. Est-ce qu'alors l'information a perdu ses effets, ou bien n'y a-t-il plus d'information?

Qu'est-ce donc que l'information?

En fait, dans les nombreux exemples précédents plusieurs caractères peuvent être distingués, qui soulignent l'originalité du concept.

Le premier télégramme, célèbre par sa concision.



des effets disproportionnés aux causes

Nous avons vu l'affichette de mobilisation, et les conséquences désastreuses qu'elle a entraînées. La capsule barométrique de la bombe de Nagasaki a « prévenu » le détonateur que la bonne altitude était atteinte. Ce dernier a fermé le circuit électrique de mise à feu de la charge de poudre... On ne peut qu'être frappé de l'extrême petitesse des dépenses énergétiques nécessaires pour l'impression, le transport et le collage des affiches ou pour la manœuvre de la capsule. **Il n'y a, en outre, aucune proportionnalité décelable entre cause et effet :** les résultats sont identiques pour une affiche deux fois plus petite ou pour une capsule dix fois plus grande.

Cette disparité entre les effets énergétiques grandioses et l'information, cause dont l'intensité est infime, fait penser à un catalyseur : on produit des tonnes d'acide sulfurique par passage de gaz sulfureux sur quelques kilos de mousse de platine, dont quelques grammes suffisent d'ailleurs pour faire détoner n'importe quel volume d'oxygène et d'hydrogène mélangés. L'information serait-elle un « catalyseur » de l'activité?

On peut le penser quand un ordre banal suffit pour précipiter de la montagne des tonnes d'eau sur une turbine électrique. Il semble pourtant que l'analogie ne doive pas être poussée trop loin. Les catalyseurs chimiques s'usent ou, comme l'on dit, « s'empoisonnent ». On connaît, par exemple, la consommation de platine par tonne d'acide produit (l'information s'use aussi, mais cela est une autre histoire).

Retenons donc ce premier aspect de l'information : **cause de niveau généralement très faible, mettant parfois en jeu d'énormes énergies.**

l'information, c'est l'inattendu

Cette remarque est absolument banale et elle correspond très bien à l'idée que nous nous faisons couramment de l'information. Un homme, une machine viennent d'être informés lorsqu'ils ont appris quelque chose de nouveau. **Si les termes du message sont déjà connus, il n'y a pas d'information.** Pas plus que n'en contient la table de multiplication une fois que l'élève la sait, ni un vieux journal déjà parcouru. Sans doute, dans la majorité des cas, une « nouvelle » n'est jamais complètement inattendue.

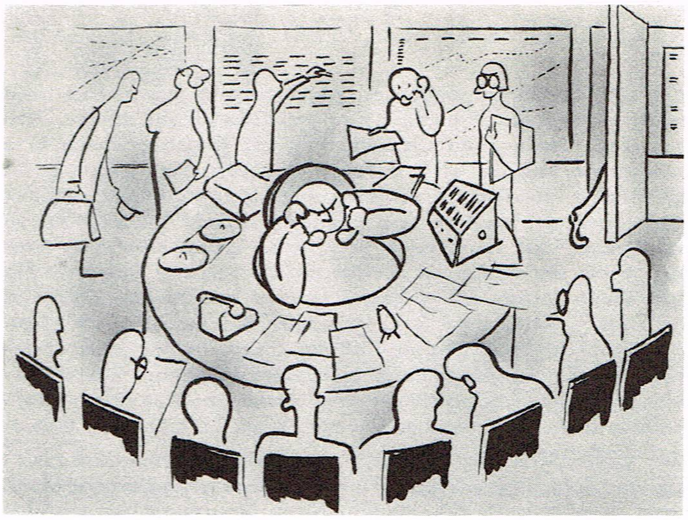
France **BAT** Angleterre **15-0**

La première partie du titre était probable en 1964; la seconde, beaucoup moins prévisible. Nous verrons plus loin à quoi correspond, de façon précise, la plus ou moins grande incertitude qui affecte l'information.

l'information, support principal des rapports entre l'homme et l'univers

On ne peut manquer d'être frappé par ce fait très remarquable : dans tous les exemples précédemment cités, où la notion d'information qui s'en dégage correspond d'ailleurs au concept courant que nous avons d'elle, l'homme est concerné, et l'est directement.

De quelque façon qu'elle intervienne, l'information aboutit toujours à l'homme, dont elle sollicite l'intelligence (que nous définissons seulement ici comme la faculté qu'il a de comprendre).



L'information aboutit toujours à l'homme.

Au sujet de cette participation de l'homme, il faut revenir à une notion à laquelle nous ne tarderons pas à donner un caractère scientifique. Ce fait est révolutionnaire. La physique traditionnelle s'est toujours vantée de sa qualité d'objectivité. Elle décrit le monde indépendamment de l'homme, autant que les lois qu'elle établit ne présupposent pas sa présence (sauf pour les établir). Or, la chaîne qui lie l'homme à l'univers comporte nécessairement trois maillons : l'homme lui-même, ses instruments de mesure et de perception et l'univers. Et cet ensemble observateur-observant-observé forme un tout que l'on ne peut dissocier.

Nous verrons combien étroitement la mécanique ondulatoire vit de la réflexion concernant l'interaction des deux derniers organes. Eh bien, la théorie de l'information ne dissocie pas davantage le premier maillon des deux autres. Serait-ce ambitieux de la définir comme une étude des rapports entre l'homme et le monde extérieur? Après avoir défini le rôle des instruments de mesure, fixer la qualité de l'intervention humaine dans la physique apparaît comme une étape nécessaire. (Certains vont peut-être crier au subjectivisme, tant il est en science comme ailleurs des mots prêts à servir pour ranimer de vieilles querelles philosophiques. L'information intéresse l'intelligence en général, mais non tel ou tel cerveau en particulier. Il n'y a pas de subjectif là-dedans.)

l'information, c'est ce qui se comprend

De paroles brouillées à la radio, du message tronqué d'un cryptogramme dont on ignore la clé, il ne sort rien : l'information a pour ainsi dire disparu. Elle doit donc être par essence significative, et l'homme doit pouvoir la comprendre.

GRAND-MÈRE DÉCIDÉE STOP REVIENS DE SUITE

GRAND-MÈRE DÉCÉDÉE STOP REVIENS DE SUITE

Un célèbre brouillage : le message n'était pas assez redondant.

Des phrases insaisissables à la radio ou des images de télévision inextricables n'impliquent pas toujours que le présentateur se soit mal exprimé ou que la caméra ait pris ses vues incorrectement. C'est peut-être en cours de route que l'information a été

détruite ou mutilée; nous verrons même qu'une déformation du message se produit nécessairement lors de sa transmission.

Pour nous résumer et mieux expliciter les quatre caractères dont nous l'avons affectée, nous pouvons proposer cette première définition de l'information : l'information est d'abord ce qui, indifféremment, peut être traduit d'une façon compréhensible par des paroles, des écrits, des images, lesquels, par un contenu nouveau et inattendu, s'adressent à l'intelligence et déterminent en général, de la part de l'homme, des réactions dont l'ampleur est sans commune mesure avec les causes énergétiques qui les ont produites.

Mais la notion ainsi dégagée retient aussitôt l'intérêt du savant. L'étude des caractères communs des choses est l'objet essentiel, sinon unique, de la science mathématique. Il ne reste qu'un petit rien pour que le concept conduise à l'érection d'une science physique de l'information : que cette dernière soit mesurable et qu'on lui applique en conséquence notre science des nombres.

Ce « petit rien » a vu le jour en 1948; il est dû au génie de Shannon. Dès lors, la notion d'information, bâtie sur un grand nombre d'idées et d'impressions souvent très anciennes que nous avions sur elle, s'est transformée en un concept de caractère général portant sur une quantité mesurable. Ainsi a pu naître une théorie scientifique de l'information.

l'information se mesure

Il vient d'être rappelé que généralité et accessibilité à la mesure sont deux caractères nécessaires du fait scientifique. C'est pourquoi l'Américain Shannon, en montrant, en 1948, comment on pouvait mesurer l'information, lui a donné ses lettres de noblesse et a construit, à partir de cette notion très ancienne et très éparse, une solide théorie scientifique.

L'exposé en apparaîtra certes très cartésien, et conforme à nos traditionnelles habitudes de pensée. Le lecteur pourra certes s'étonner alors que des lois si nécessaires aient pu être dégagées si tard : c'est qu'il se trompe souvent sur la nature de la recherche scientifique et de la découverte. La méthode, qui met l'ordre et la clarté dans l'acquis, cède bien souvent le pas à l'intuition, à la patience, voire à l'audace — toutes qualités humaines nécessaires lorsque, au contact de l'inconnu, l'homme utilise mal ses outils classiques. Il reste que l'essentiel peut être compris et développé à partir des exemples les plus simples.

l'unité d'information

La manière la plus simple pour transmettre une information est de n'utiliser que deux types de caractères : suite de traits ou points de l'écriture morse, relevé des sorties du rouge ou du noir à la roulette (R N).

Vous venez de recevoir un message de 20 caractères de ce type. Avant de les avoir lus, vous ignorez tout de son contenu; il y a autant de chances que le premier signe soit donc un trait ou un point (R ou N). Il n'est donc pas absurde de dire que la lecture du premier symbole vous a communiqué une unité d'information. Et une deuxième la lecture du deuxième (supposé indépendant du premier), et ainsi de suite.

Or, combien de formes différentes peuvent revêtir les messages reçus? Chaque signe peut être indifféremment un trait ou un point, le rouge ou le noir. Au total, il y a $N = 2^{20}$ formes différentes possibles. D'après la définition, l'information I contenue dans le message était de 20 unités. On a donc

$$I = \log_2 N \text{ (1)}.$$

A partir d'une définition extrêmement simple de la quantité d'information, nous constatons que nous sommes capables d'atteindre une mesure beaucoup plus générale et puissante de l'information contenue dans un message. Si un message comporte *a priori* N formes possibles (et équiprobables), l'information qu'il apporte est mesurée par

$$\mathbf{I = \log_2 N.}$$

L'unité a reçu le nom de « logon » (ou, encore, de « bit », abréviation de « binary unit », que l'on pourrait traduire littéralement par « ubi », unité binaire d'information).

une écriture nécessairement redondante

Qu'en est-il de l'alphabet et de l'écriture en général? Imaginons que le même correspondant qui a adressé les traits en morse et les coups de roulette envoie maintenant un message formé de cinq lettres d'alphabet absolument quelconques. Le nouveau message peut revêtir 26^5 formes différentes, et, comme tel, il porte l'information

$$I = \log_2 26^5 \sim 23 \text{ unités.}$$

On constate en passant qu'il y a davantage d'information dans une suite de cinq lettres que dans un nombre de six chiffres.

Mais votre partenaire aurait pu aussi ne vous envoyer que des mots français de cinq lettres, choisis au hasard. Il y en a environ 5 000, et l'information dont il vous eût gratifié eût été de $\log_2 5\,000$, soit douze logons seulement. Cela s'explique très bien parce que, parmi toutes les combinaisons possibles de cinq lettres, il y en a réellement assez peu qui correspondent à un mot français. On s'aperçoit d'ailleurs que trois lettres auraient suffi pour signifier ces 5 000 mots puisque l'on peut les combiner de $26^3 \sim 16\,000$ manières différentes. On dit que le langage écrit est **redondant,** c'est-à-dire que l'association des lettres qui servent à le former est loin d'épuiser toutes les possibilités, ou — ce qui revient au même — qu'on pourrait le contracter et réduire par là le nombre des signes nécessaires. Par exemple, le dictionnaire français comporte quelque 35 000 mots, de toute longueur. On peut former 26^4 (450 000) mots quelconques de quatre lettres. On pourrait donc réduire la langue française à 35 000 mots de quatre lettres au plus, dont d'ailleurs 16 000 (26^3) de trois lettres. Ce qui n'irait d'ailleurs pas sans dommage pour la prononciation et pour la poésie.

vers une concision de plus en plus grande

Pourtant, de telles contractions sont couramment utilisées dans la foule des abréviations et dans le langage dit « télégraphique ».

**J'ARRIVERAI DEMAIN LE QUATORZE JUILLET
A LA GARE MONTPARNASSE
A SEIZE HEURES QUARANTE-CINQ**

(1) De même que le logarithme décimal d'un nombre N est le nombre x tel que $10^x = N$, le logarithme à base 2 de N est le nombre y tel que $2^y = N$. Ainsi, $\log_2 8 = 3$ ($8 = 2^3$), $\log_2 64 = 6$ ($64 = 2^6$), etc.

Ce télégramme de 75 lettres, qui a été rédigé par un homme riche, contient $\log_2 26^{75} = 350$ unités d'information. Mais on se doute bien qu'il est trop largement explicité, c'est-à-dire très redondant. On commencera par écrire les dates et heures avec des chiffres, afin de supprimer « demain », qui n'apporte aucun renseignement supplémentaire puisque le télégramme est daté, pour obtenir un énoncé plus court, mais toujours « français » :

J'ARRIVERAI LE 14 JUILLET A LA GARE MONTPARNASSE A 16 HEURES 45

Au total 45 lettres et 6 chiffres. Pour ces derniers, il faut remarquer qu'il n'y a que 31 jours dans le mois de juillet, 24 heures par jour et 60 minutes dans l'heure. L'information de $\log_2 (26^{45} \times 31 \times 24 \times 60)$ est 212 bits.

Mais nous savons déjà que le français est très redondant.

ARR 14 7 GARE MTPNSSE 16 45

restera un libellé parfaitement compréhensible : 14 lettres, 7 chiffres.

$$I = \log_2 (26^{14} \times 31 \times 12 \times 24 \times 60) = 140 \text{ logons.}$$

Mais on peut faire mieux. Sans aller jusqu'à supposer que l'invité, venant de Bretagne, est forcé d'arriver à la gare du Montparnasse, qu'il doit atteindre Paris par un train de la soirée du 14 et que seule subsiste l'incertitude entre trois trains, auquel cas, si son choix se portait sur le deuxième, il aurait pu adresser le télégramme

2

ne nécessitant que $\log_2 3 = 1,6$ unité d'information, on peut remarquer :

— qu'il n'y a que six gares principales à Paris;
— que l'invité arrivera sûrement en juillet;
— que, se dirigeant sur Paris, il ne peut qu'y arriver.

Le destinataire comprendrait encore le télégramme (18 logons) :

14 MTPNSSE 16 45

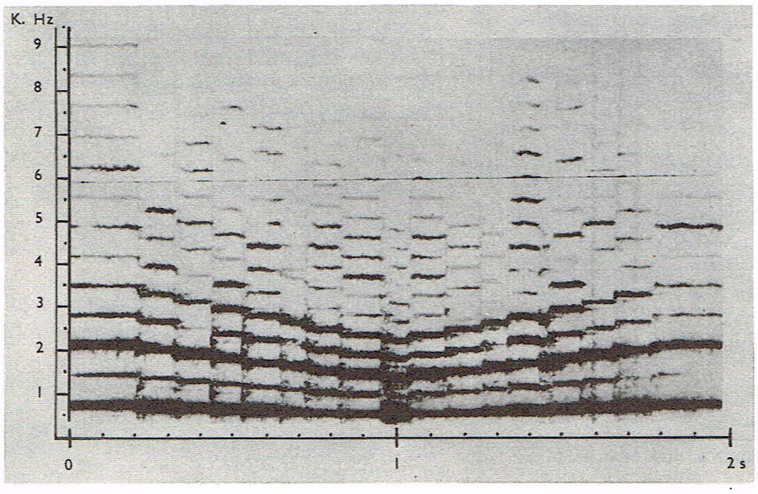
La redondance du message, qui s'exprime par la différence entre les 350 bits d'information apparente et les 18 d'information réelle, est énorme. C'est ce qu'on nomme « verbiage » en littérature. Il va de soi que c'est le lot des romans-fleuves, des articles de journaux, des déclarations électorales et des discours de plus d'un chef. C'est ce que le public exprime quand il constate que l'orateur « n'a rien dit de nouveau » (information zéro).

On sait, par ailleurs, qu'une image de télévision comprend en France 819 lignes, chacune d'elles étant formée d'environ 760 points évoluant entre le noir absolu et le blanc absolu. En conséquence, elle peut comporter (dans le cas simplifié de l'absence de demi-teinte) jusqu'à $760 \times 819 = 622\,440$ cases noires ou blanches. Chaque case ne possédant dans ce cas que deux états possibles, l'information transportée alors par une image de télévision est de 622 440 unités, soit $620\,440 \times 25 \sim 1,55.10^7$ unités/seconde (25 images à la seconde). C'est un total considérable, et l'on devine que, là aussi, la redondance doive être considérable. Son calcul cependant est très difficile et dépend évidemment de la nature des images projetées. Il va de soi que le visage humain n'est pas quelconque et que telle succession de points noirs qui amorcent le dessin de la ligne du nez a son évolution définie dans son ensemble. C'est ce que savent très bien les dessinateurs et caricaturistes, dont l'art s'applique à réduire au minimum la redondance de leurs croquis par un trait ramené à l'essentiel.

les règles de la quantification permettent de coder tout message, fût-il sensoriel

Nous venons de voir comment coder un message aussi complexe et d'apparence aussi continue que celui de telle émission de télévision ou de tel panorama. Dans ce dernier cas n'y a-t-il pas une difficulté supplémentaire : comment coder la couleur? Physiciens et physiologistes sont d'accord pour donner la même réponse : notre œil filtre le message lumineux, qui impressionne des cellules (cônes rétiniens) sensibles à une seule gamme de longueurs d'onde. Toute couleur se traduit donc par les valeurs des trois intensités reçues : en bleu, en rouge et en jaune. On sait que sur le même principe de la superposition de trois couleurs de base sont fondées l'échelle internationale des couleurs, la photographie ou la télévision en couleurs, et l'impression en trichromie.

Un message sonore peut se décomposer de même; il suffit d'indiquer la variation en fonction du temps de l'intensité reçue sur les diverses longueurs d'onde qui traduisent une sensation perçue par l'oreille. Là encore le problème de codification semble insoluble, puisqu'il faudrait une infinité de courbes, donc une double infinité de valeurs, pour connaître l'intensité $I(t, \lambda)$ à chaque moment et à chaque période. Là encore la réponse est pourtant simple si l'on tient compte de la nécessaire quantification : il suffit de connaître les valeurs de I qui correspondent à celles des longueurs d'onde que l'oreille enregistre distinctement, et l'intensité moyenne au cours du quantum de temps au-dessous duquel l'oreille ne peut descendre par suite de l'inertie des organes mécaniques que sont ses osselets.



Même le son peut se décomposer, et la figure ci-contre est le sonogramme d'une ritournelle de clarinette : on indique l'intensité reçue chaque fraction de seconde sur les diverses longueurs de fréquences. La ligne inférieure est la ligne de base (85 Hz) ; les lignes suivantes, en montant, sont les harmoniques. ***(Doc. E. Leipp.)***

Phot. Larousse.

On ne sait encore exactement quels sont les codages successifs que revêtent les messages sonores ou visuels : de gros efforts sont consacrés à de telles études. Mais il est certain qu'il s'agit de systèmes analogues à celui que traduit le sonogramme, et que la notion de filtrage joue ici aussi un rôle essentiel.

Peut-on aussi bien coder un message tel qu'un enregistrement? Quantification et codage étant liés, peut-on quantifier ici aussi?

Une première réponse s'impose : toute connaissance ne pouvant nous être transmise que par un organe sensoriel, il faut bien lire — un jour ou l'autre — l'enregistrement, ce qui nous ramène au cas du paragraphe précédent. **Mais il faut aller plus loin : tout système physique — et non seulement l'homme — obéit aux lois de la quantification, l'être vivant ne présentant aucune originalité à ce point de vue (ni d'ailleurs à beaucoup d'autres).**

Considérons la courbe qui donne la réponse d'un récepteur à la réception d'un message : il y a un seuil au-dessous duquel l'appareil reste au repos, et un autre au-delà duquel il se dérègle ou ne répond plus que d'une manière uniforme (seuil de saturation). De plus, deux messages trop voisins ne sont pas discernés (seuil différentiel de sensibilité). Finalement, la réponse ne peut prendre qu'un nombre fini et bien déterminé de valeurs. On retrouve la loi de quantification propre à tout système fini (le nombre d'états peut être très grand, il sera toujours limité : une bascule pouvant peser de 1 à 100 kg et graduée de 100 en 100 grammes peut prendre 990 états discernables; on peut améliorer la sensibilité et l'amener à distinguer 10 grammes; le nombre d'états discernables passe à 9 900, mais il n'en reste pas moins fini).

l'information ne peut se transmettre qu'en se dégradant

Nous nous sommes posé le problème de la redondance des messages. Cette recherche de l'économie d'expression et de moyens, la théorie de l'information l'a partagée avec les arts graphiques. Et même avec les mathématiques, qui apparaissent comme un langage très condensé, très économique, et, partant, très peu redondant. Peut-être même est-ce là que réside l'unique secret de leur puissance, l'homme étant parvenu à se construire une représentation du monde débarrassée de la lourde redondance de la langue courante.

Ce progrès fut très visible au cours des siècles : il n'y a qu'à se reporter aux interminables démonstrations de nos ancêtres mathématiciens et à les comparer au symbolisme actuel.

On pourrait être surpris du souci, qui semble transparaître depuis quelques pages, de la concision nécessaire du langage, de l'écriture, des images. Ce scrupule participe, bien sûr, d'un sentiment esthétique, celui-là même qui constate que l'œuvre concise est généralement belle. La raison en est peut-être même purement biologique : il s'agirait d'une sorte d'avarice intellectuelle.

Mais nous sommes physiciens et contribuables; or, l'information possède ce défaut de coûter cher au stockage et à la transmission. Et l'on comprend aisément que ce soit là un motif solide pour justifier tous les efforts faits en vue de la réduire à son expression suffisante.

l'inévitable bruit de fond

Que faire d'une information? Sinon la comparer à d'autres, l'emmagasiner dans notre cerveau ou, sous forme d'un document imprimé, la communiquer. En un mot, la transmettre.

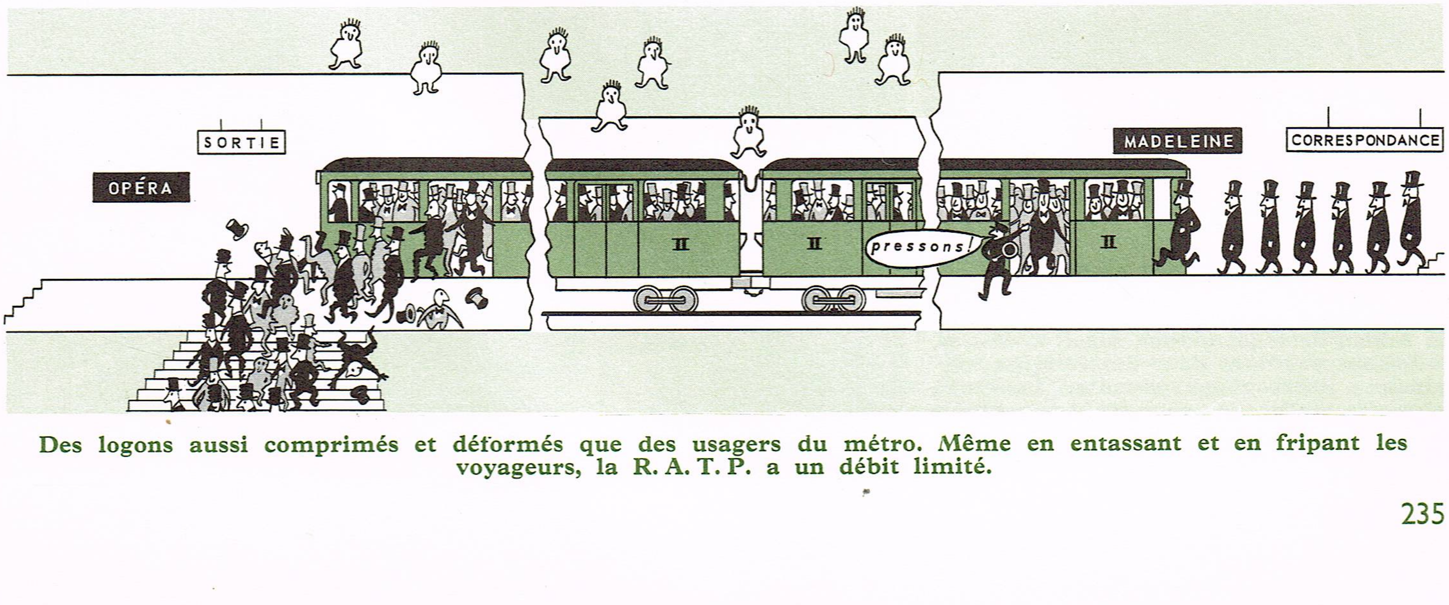
Nous avons déjà vu, dans le tome I, que ce seul fait entraîne deux limitations :

— On la déforme; que ce soit sous l'effet des parasites, qui s'ajoutent inévitablement au message et qui le brouillent, ou de ce reflet de l'agitation thermique des électrons, qui se traduit par un bruit de fond encore plus fondamentalement inéluctable, le message est déformé et finalement appauvri;

— Le débit en est limité. Tout système physique se comporte comme un filtre et ne laisse que certaines des composantes périodiques en lesquelles se décompose nécessairement un message. Concevoir que cela se traduit par une limitation du débit d'information de la forme

$$C \leqslant L \operatorname{Log} (1 + S/B)$$

(formule due à Shannon; C est le débit en bits/seconde; L, la largeur de bande définissant le canal; S/B, le rapport des intensités du signal à transmettre et des bruits affectant la transmission) n'est peut-être pas immédiat, mais le fait n'en est pas moins vrai : **une ligne de téléphone ne peut transmettre que 50 000 bits/s et un canal de télévision 100 millions.**



Des logons aussi comprimés et déformés que des usagers du métro. Même en entassant et en fripant les voyageurs, la R. A. T. P. a un débit limité.

le problème des télécommunications

On peut poser le problème de la communication de l'information en ces termes précis : « permettre à un homme ou à une machine situés en un lieu A d'échanger un message avec leurs homologues situés en B, et ce de manière aussi efficace et peu coûteuse que possible (là se posent les problèmes du codage optimal et de la redondance minimale compatible avec la détection des erreurs les plus habituelles) ».

l'acquisition de l'information

La première phase est dite « acquisition ». Elle consiste à détecter, c'est-à-dire à mettre sous une forme pratique l'information que l'abonné A désire expédier. Cette opération, qui s'effectue en tête de la chaîne des télécommunications, s'opère avec le microphone quand elle concerne la parole, le téléimprimeur pour l'écriture, l'appareil de photo, la caméra de télévision, les détecteurs de température, de pression, de champ magnétique, le ruban magnétique, le perforateur de cartes, la machine à calculer lorsqu'elle concerne télévision ou données.

la transmission de l'information

Une fois l'information acquise, il faut la transporter. C'est l'opération « transmission », qui est d'ailleurs généralement résolue, à la satisfaction de tous (pour le courrier), par les trains, les avions, les camions et les facteurs. L'électricité, toutefois, offre un véhicule beaucoup plus rapide (il atteint 300 000 km/s). On utilisera en conséquence les milieux les mieux adaptés à la propagation du champ électrique, ou, mieux, du champ électromagnétique.

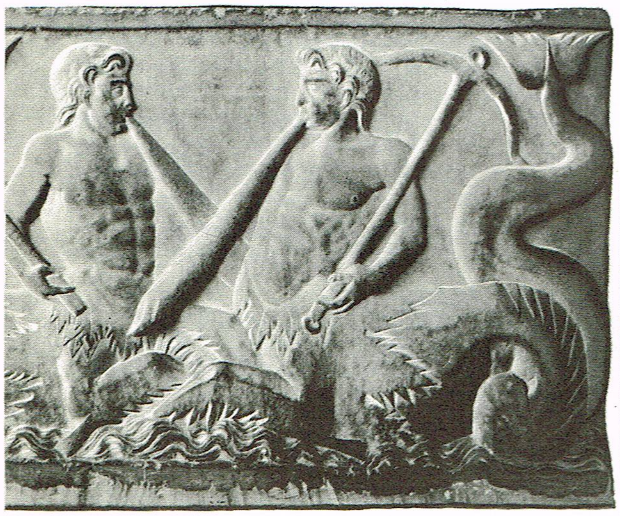
On distingue volontiers les milieux naturels (l'air, le vide, l'eau, la terre) des milieux artificiels ou métalliques (fils, câbles coaxiaux, guides d'ondes). On demandera évidemment à ces milieux d'être peu dispersifs, c'est-à-dire de ne pas éparpiller l'énergie qu'on leur fournit, et peu dissipatifs, c'est-à-dire de ne l'absorber que modérément.

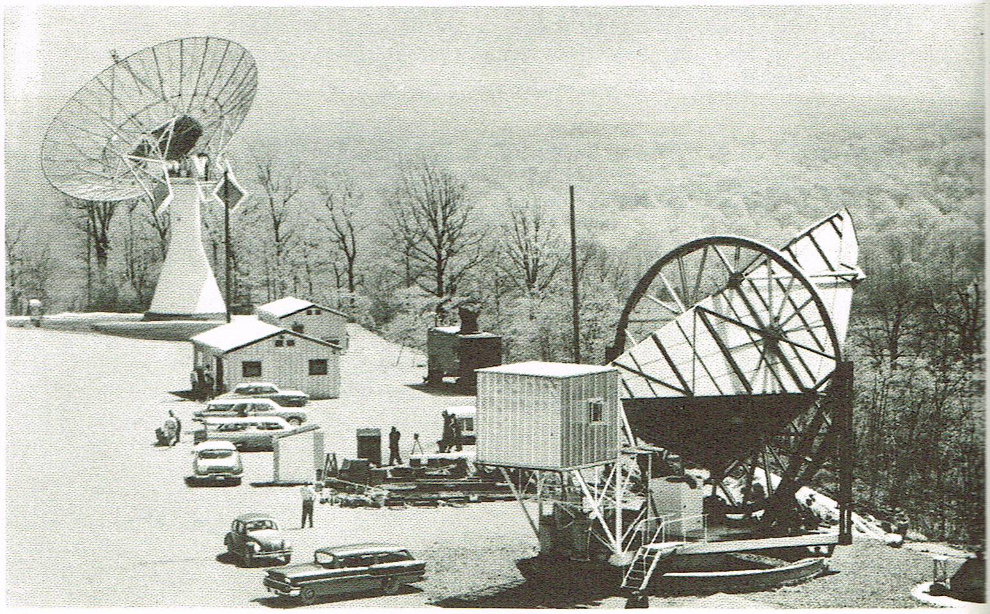
D'où il résulte que, sans nul doute, le vide est le milieu de transmission idéal (satellites et communications extra-terrestres l'utilisent). Ainsi des liaisons Terre-Mars et Terre-Vénus ont-elles pu être réalisées dans des conditions excellentes et spectaculaires. Le vide possède la propriété de transmettre pratiquement sans perte toutes les fréquences, depuis les ondes très longues jusqu'aux ondes ultra-courtes (rayons X, rayons γ, rayons cosmiques). Il n'a qu'un inconvénient, que les astronomes tiennent d'ailleurs pour un avantage remarquable : il est « bruyant », l'espace renfermant un grand nombre de parasites inévitables (Soleil, émettant principalement dans le spectre visible; galaxies; autres radiosources [v. le chapitre consacré à la radio-astronomie]).

Pour nous qui restons sur la planète, l'air est, à défaut du vide, le domaine privilégié des ondes hertziennes. Les limitations qu'il nous impose sont très supportables. Il ne transmet malheureusement que les longueurs d'onde supérieures à 1 cm (absorption due à la vapeur d'eau et au gaz carbonique, puis à l'oxygène et à l'azote), et, quand on prend les longueurs d'onde décroissantes, redevient translucide dans la bande des fréquences visibles (0,4 à 0,8 μ).

Toutes ces qualités s'évanouissent, au grand dam des sous-mariniers, lorsqu'on utilise l'eau. Tout au plus, les ondes très longues peuvent-elles pénétrer quelque peu dans les profondeurs océanes (le sceau du secret frappe tous les procédés de communications sous-marines, qui donnent lieu à de très importants travaux).

Enfin, la terre, chacun le sait, est un milieu très absorbant. En 1914, quelques essais furent tentés avec plus ou moins de succès dans le domaine des communications souterraines entre tranchées. On parlait alors de T. P. S. (transmissions par le sol). Ces tentatives ont été reprises de nos jours, et notamment aux États-Unis, en mettant en œuvre les procédés les plus modernes. Il semble que de solides espoirs puissent être fondés sur la structure géologique des sols. Il est notamment fort possible que, dans les couches profondes (roches anciennes, granite principalement), on arrive à produire une propagation guidée des ondes de la façon suivante : l'énergie dégagée dans le sein de la masse granitique resterait canalisée grâce à l'existence des couches superficielles plus conductrices (nappes phréatiques,





La pression de l'air était autrefois le seul état physique que l'on sût modeler, mais le vide nous apparaît aujourd'hui comme le milieu de transmission idéal. *Ci-dessus*, « Tritons soufflant dans des conques marines » (sarcophage romain, Louvre) ; *ci-contre*, station de poursuite et de contrôle des engins spatiaux.

Phot. Giraudon, U. S. I. S.

roches sédimentaires) aussi bien que celle des couches très profondes plus chaudes, donc également plus conductrices. On éviterait ainsi un éparpillement des ondes en limitant de la sorte l'absorption toujours considérable qu'elles subissent.

Les exigences de la non-dissipativité et de la non-dispersivité sont correctement remplies par les milieux artificiels métalliques, parmi lesquels il faut citer : les fils tels qu'on les voit sur nos poteaux télégraphiques, les coaxiaux, qui sont constitués par un conducteur central séparé d'une enveloppe métallique cylindrique par un isolant (polyéthylène, Téflon) et qui s'adaptent bien aux fortes capacités; enfin les guides d'ondes, conducteurs creux généralement rectangulaires, qui fonctionnent pratiquement sur une fréquence déterminée et servent généralement aux liaisons (courtes) entre les émetteurs-récepteurs de faisceaux hertziens, les radars et les antennes.

la modulation

Quoi qu'il en soit, et une fois le milieu choisi, il faut savoir lui adapter au mieux l'information que l'on désire lui faire transporter. Le procédé que l'on utilise s'appelle la « modulation ». **Moduler, c'est faire varier le champ électromagnétique selon l'information qu'on lui applique.**

Or, les procédés possibles de modulation sont peu nombreux, et, pratiquement, il n'en existe que trois.

Le plus courant consiste à faire dépendre proportionnellement l'amplitude du champ de l'amplitude de l'information. C'est précisément la « modulation d'amplitude ». Ainsi Radio-Luxembourg utilise-t-il une longueur d'onde de 1 293 m, ce qui veut dire que sa fréquence d'émission est de 232 kilocycles. Quant à l'amplitude de cette dernière fréquence, elle dépend uniquement de l'intensité de la voix ou du fond sonore qui lui est appliqué. On montre d'ailleurs dans ce procédé que, si le présentateur émet le la_3 (435 vibrations par seconde), l'émetteur émet la fréquence porteuse, 232 kilocycles, et les deux fréquences « latérales », 232 kHz + 435 Hz et 232 kHz — 435 Hz. Ainsi, pour un concert où les fréquences s'étalent de 50 à 10 000 Hz (0,050 à 10 kilocycles/seconde), Radio-Luxembourg envoie dans l'espace la fréquence porteuse à 232 kHz, une bande latérale inférieure allant de 222 (232—10) à 231,950 (232—0,050), une bande latérale supérieure de 232,050 (232 + 0,050) à 242 (232 + 10) kHz.

Un autre procédé très apprécié des mélomanes est la « modulation de fréquence », utilisée par certains postes (France Musique) et les faisceaux hertziens en général. A chaque instant, la fréquence émise par le poste dépend de l'information qui lui est appliquée, tandis que l'amplitude reste constante. L'avantage est le suivant : tant que le signal est apparemment supérieur au bruit de fond parasite, la qualité du son est excellente et bien supérieure à celle que permet disons 10 fois la modulation d'amplitude.

Enfin une méthode toute nouvelle et très prometteuse commence à être employée : celle de la « modulation codée ». Un signal téléphonique, par exemple, se présente toujours comme une intensité de courant, ou une tension variant en fonction du temps. De plus, son spectre a une étendue finie : de 300 à 3 400 Hz pour le téléphone standard, de 20 à 15 000 Hz pour la haute fidélité des électrophones. On montre alors que l'on ne modifie en rien l'information en prélevant de courts échantillons du message à transmettre et en reconstituant le message à partir d'eux. Ainsi

$2 \times (3\,400 - 300) = 6\,200$ impulsions par seconde suffisent pour décomposer un message téléphonique.

Si l'on accepte d'utiliser 64 niveaux (ce qui signifie que l'on remplace la hauteur h exacte par un multiple nh_0, h_0 étant le seuil, et n, un entier variant de 0 à 63), il suffit de coder n au moyen de 6 signaux ($2^6 = 64$).

Phot. C. S. F.



Réflecteur parabolique de 16 m de diamètre, utilisé pour des liaisons troposphériques.
(Doc. C. S. F.)

Finalement, on utilise $6 \times 6\,200 = 37\,200$ impulsions pour décomposer, transmettre et recomposer un message. Il n'est pas étonnant qu'un codage si complexe (procédé P. C. M. ou « pulse code modulation ») constitue une efficace protection contre le bruit de fond.

Les appareils de transmission courants sont : les micros téléphoniques, qui envoient directement la parole sur les fils; les téléimprimeurs, qui codent les lettres et signes sous la forme de cinq impulsions de courant; les émetteurs de radiodiffusion, qui emploient depuis les ondes longues (1 km) jusqu'aux ondes courtes (quelques décamètres), ces dernières possédant la propriété d'être réfléchies par les couches ionisées de la stratosphère, ce qui permet de très longues portées; les émetteurs dits « V. H. F. » et « U. H. F. » (longueur d'onde inférieure à 10 m), dont la fréquence élevée permet d'utiliser des antennes de plus en plus directives. Ces émetteurs sont utilisés dans les liaisons sol-air et air-sol; leur portée est cependant limitée (pratiquement réduite à l'horizon), car les ondes qu'ils émettent ne sont pas susceptibles de se réfléchir sur l'ionosphère.

le traitement de l'information

Cette opération est commune à toutes les transformations que subit l'information, à l'exclusion de celle qui consiste à l'appliquer au support de transmission (modulation). On a vu le multiplexage à courants porteurs, permettant, grâce au décalage des spectres, de juxtaposer un grand nombre de voies (téléphoniques, télégraphiques) avant de les faire débiter par raison d'économie sur un même support.

On peut, à cette occasion, citer un autre procédé, encore récent, de multiplexage. Comme l'information circulant sur chaque canal peut être représentée par une succession d'impulsions, on peut réduire la longueur de chacune d'elles de façon à imbriquer dans le temps les impulsions appartenant à des canaux différents : c'est le multiplexage par division dans le temps, qui a fait, lui aussi, l'objet d'un paragraphe spécial du premier tome de cet ouvrage.

Un autre traitement, d'une importance capitale, concerne la distribution de l'information. Elle est effectuée dans des ensembles appelés « commutateurs », dont les centraux des P. T. T. sont les illustrations les plus typiques.

Enfin, la technique moderne s'intéresse non seulement au multiplexage et au routage, mais aussi aux opérations qui modifient profondément l'information elle-même. C'est le rôle des calculateurs de toute sorte, depuis la machine de bureau jusqu'aux énormes calculatrices universelles. (V. tome Ier, page 205.) Après ces modifications, l'information sort ordonnée, condensée ou explicitée, voire même résolue (calculs proprement dits).

la présentation de l'information

Après l'acquisition, la transmission, un éventuel traitement, l'information doit être présentée commodément à l'abonné utilisateur. On pourra remarquer que présentation et acquisition, qui sont aux deux bouts de la chaîne des télécommunications, sont deux opérations symétriques. Il suffit d'ailleurs de citer les équipements concernés pour le constater :

— l'écouteur et le haut-parleur du téléphone;

— le téléimprimeur pour le télégraphe;

— le tube cathodique de la télévision, l'écran radar, les projecteurs divers, les visualisateurs, les perforateurs de cartes, les enregistreurs sur papier, sur bandes magnétiques, pour les images et les données...

signification et valeur humaine de l'information

Il est peut-être temps de faire amende honorable et de reconnaître que beaucoup d'imprécisions ont été laissées dans l'exposé de la théorie. En fait, la signification que nous avons donnée implicitement au mot « information » dépasse souvent celle que lui assignent rigoureusement les spécialistes.

C'est une difficulté que Léon Brillouin définit parfaitement quand il déclare que, au regard de la théorie physique, il y a tout autant d'information dans tel document ultra-secret que dans un passage de *France-Soir* contenant le même nombre de caractères typographiques : le coût de la transmission exacte est effectivement le même pour les deux messages.

De même, le décès du Français immatriculé 2.23.01.75.112.027 mobilise, quand on l'apprend, autant d'information que celui du numéro 1.28.03.75.115.292, quand bien même ce dernier nous serait un être cher, et le premier, un anonyme.

Notre façon de mesurer l'information est donc objective, sans qu'il puisse être tenu compte de l'importance que nous attachons personnellement à telle ou telle nouvelle. D'ailleurs, c'est précisément ce caractère d'objectivité qui rend scientifique la notion définie et mesurée.

Cependant, cette théorie est un trop magnifique instrument d'approche dans l'étude des interventions de l'homme à l'intérieur du monde physique pour que ses applications soient limitées par un trop strict corset d'objectivisme. Il faudrait pouvoir assigner une valeur à tout document, comme le remarque Léon Brillouin lui-même; et il va de soi que celle du document secret (sauf pour le rédacteur même du document) doit dépasser celle de la « une » du journal vespéral.

De très nombreux travaux sont entrepris dans cette direction. On conçoit que beaucoup de facultés de l'homme vont intervenir, et notamment sa mémoire. C'est parce que son rédacteur « sait » le contenu du dossier secret, qu'il ne contient guère plus d'information pour lui que l'article de *France-Soir*.

Nous verrons bientôt le rôle que cette notion d'enrichissement de la mémoire joue en biologie : c'est la nécessité d'acquérir de l'information qui amène finalement l'homme à vivre du passé au futur (principe de Brillouin que nous retrouverons bientôt).

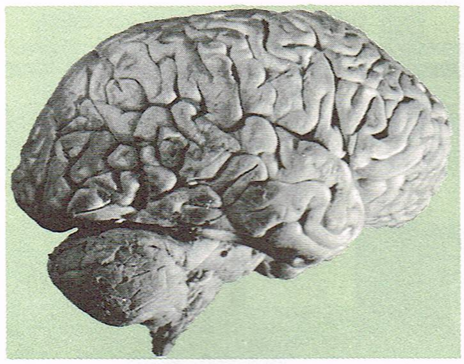
Notons, du reste, que cette mémoire n'est pas tout entière enfermée dans les molécules cérébrales; n'oublions pas l'information génétique contenue dans les molécules géantes de l'A. D. N. Nous avons vu dans le premier tome de cet ouvrage que l'étude du codage qu'elles réalisent donnerait des clés essentielles de la biologie moderne; les voies ainsi tracées ouvrent des perspectives immenses — des plus heureuses aux plus dangereuses — concernant l'avenir même du genre humain.

Reste un fait essentiel : la mémoire est limitée, même si l'on tient compte des mémoires auxiliaires que sont les livres, les microfilms, les bandes de magnétophone, etc. (la « maison de la Radio » ne peut-elle accueillir des millions de documents, qu'il faudrait des siècles pour lire tous !). *Il y a là une limite imposée à l'homme; la tâche du savant désirant synthétiser (et pour cela connaître préalablement) les travaux des spécialistes sera de plus en plus difficile, peut-être même définitivement impossible demain à un certain niveau.* A moins, précisément, que nous ne puissions élaguer convenablement notre mémoire, encombrée de tant d'informations inutiles; ou à moins encore qu'une meilleure utilisation de nos facultés, qu'un stockage plus efficace de l'information (nous avons vu au tome I que ce problème est, lui aussi, crucial pour l'avenir des ordinateurs), qu'une diminution de l'énorme redondance des renseignements que nous rangeons dans les cases de notre cerveau ne nous permettent de résoudre ce problème. **Une telle ambition n'est pas le moindre intérêt d'une discipline aussi jeune que celle que nous venons de considérer; et le secours déjà irremplaçable que nous apportent les centres de documentation automatique est là pour montrer les premières réalisations, extrêmement prometteuses.**

●

Que faire encore?

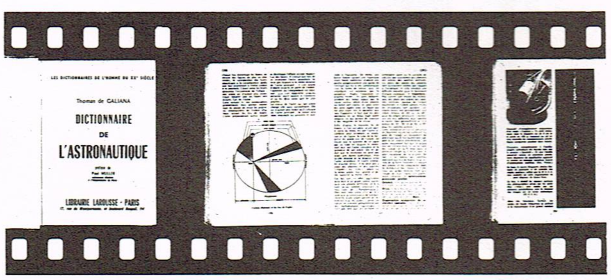
Eh bien, tirer le parti optimal de nos connaissances, qui vont en augmentant sans cesse, au lieu de nous complaire dans une contemplation morbide et stérile de notre ignorance. « Estimer et agir au mieux en fonction de notre savoir » — tel est le but que nous proposent les statisticiens. Imitons leur réalisme et écoutons-les.



Le problème du stockage le plus efficace de l'information dans le cerveau humain n'est pas résolu ; que de redondances et d'informations inutiles !



Salle de travail des imprimés à la Bibliothèque nationale.



Une bibliothèque sur microfilms.

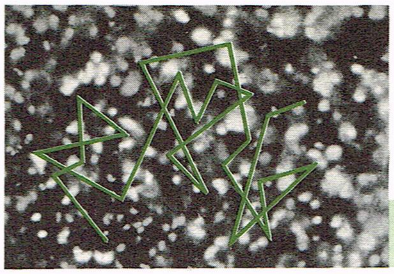


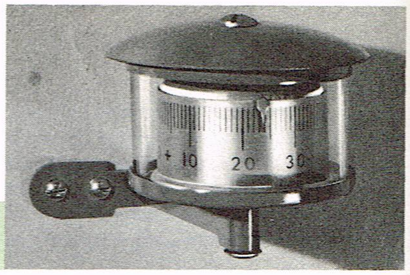
Une discothèque.

Les techniques modernes (microphotographie, discographie) permettent de diminuer considérablement le volume nécessaire au stockage de l'information. Tout cela représente une somme fantastique d'informations, mais d'un intérêt évidemment inégal.

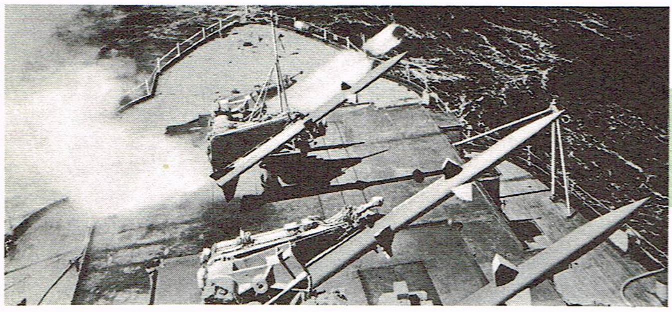
Phot. Larousse, Edimeta.

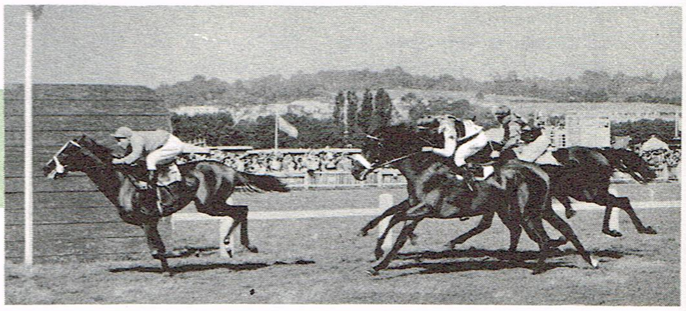
**Des faits multiples et divers,
mais un seul schéma d'explication.**





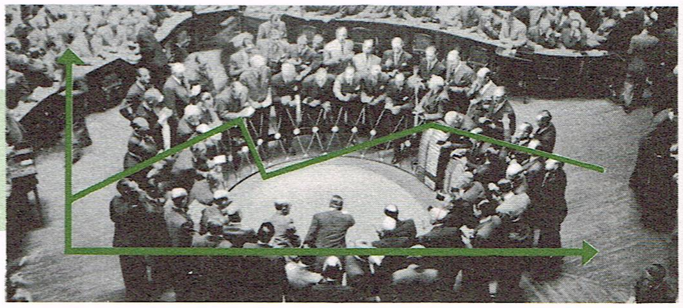
Le mouvement brownien et le thermomètre qui mesure l'agitation des particules.



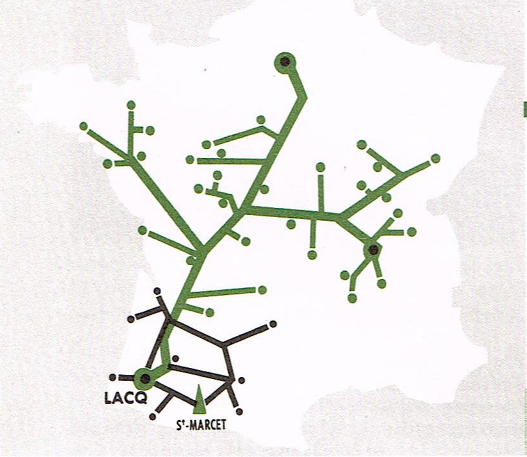


les jeux
de la guerre
et
de la paix





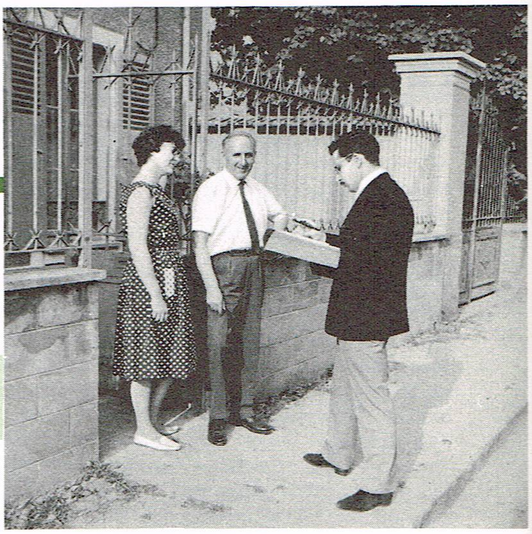
grâce à
la statistique,
des règles
efficaces
pour l'action



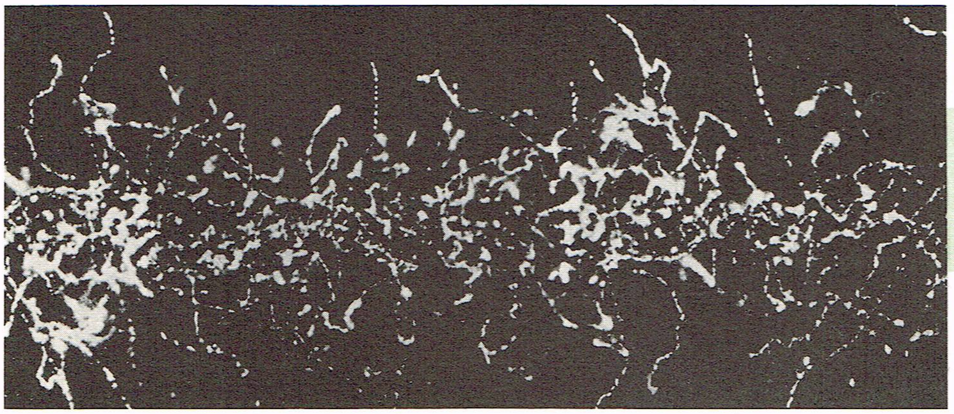
distribution
du gaz de Lacq .

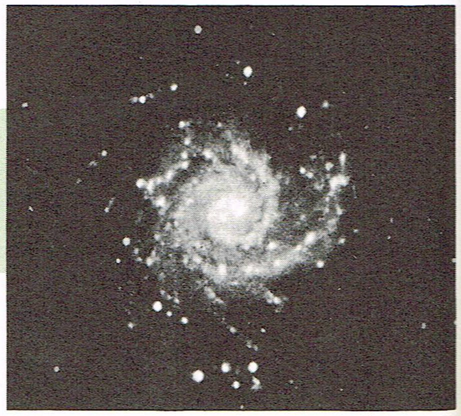
recherche
opérationnelle





du comportement
intime des chromosomes
au mode de vie





de la particule
élémentaire
à
l'organisation
de
l'Univers

Phot. Larousse, U. S. I. S., Bertrand, A. F. P., Lejeune, observatoire Lick.

STATISTIQUES et PROBABILITÉS

par

Jean-Louis RIGAL

Qui ne risque rien n'a rien

Savoir pour prévoir, prévoir pour pourvoir

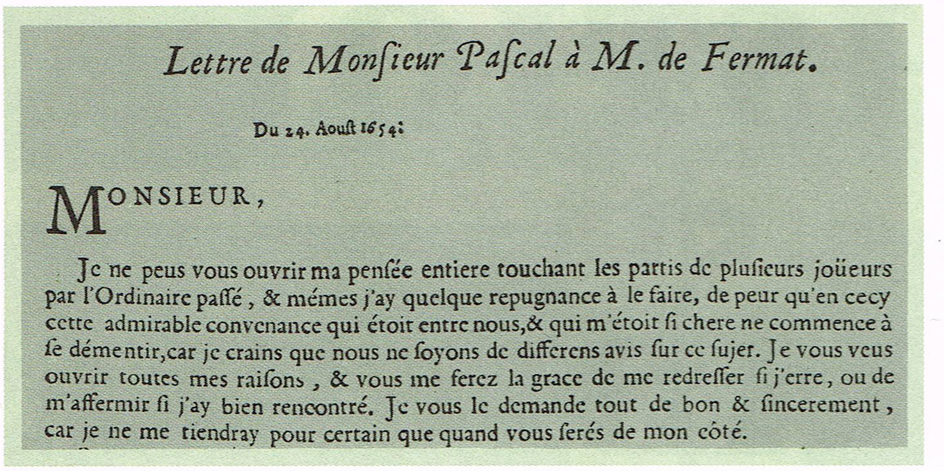
Dans l'esprit de certains, le calcul des probabilités dresse un constat de faillite chiffré de l'outil scientifique, puisqu'il remplace les lois dites « rigoureuses », auxquelles nous sommes habitués, par de vagues relations où un phénomène est décrit comme ayant un certain nombre de chances de se produire.

Pour d'autres, c'est une sorte de jonglerie arithmétique qui commence par prétendre montrer qu'on ne peut gagner ni à la roulette ni au P. M. U. — ce qui est infirmé chaque jour par l'expérience —, pour finir par présenter toute situation sous forme de statistiques, qui sont, paraît-il, la forme la plus moderne du mensonge.

Il faut reconnaître que ses fondateurs ont d'abord considéré les probabilités comme un divertissement. D'ailleurs à double titre, puisque pratiqué comme tel par des gens aussi austères que Pascal, et appliqué aux combinaisons des jeux.

Cette qualité particulière des probabilités de concerner des sujets aussi prosaïques que les jeux de hasard ne s'est pas démentie depuis trois siècles. Mais, en outre, la science probabiliste fait appel aujourd'hui à une mathématique, sans doute assez difficile, pour régler les bénéfices des compagnies d'assurance, dicter les bulletins de la météorologie, régner sur la recherche opérationnelle, c'est-à-dire sur l'art de prendre des décisions si possible motivées, commander à l'évolution des espèces biologiques, présider aux mouvements des électrons, atomes et molécules, régir la thermodynamique, empêcher les piles atomiques d'exploser, donner leur chance aux astronautes de n'être pas tués par les météorites.

Cette énumération n'aurait pas de fin. Signifie-t-elle que le calcul des probabilités est devenu puissant au point de gouverner, sous une même loi, physique, chimie, mécanique, biologie, et jusqu'à la sociologie? Elle le signifie effectivement : le monde du XXe siècle est probabiliste. Y a-t-il donc du jeu dans la machine scientifique?



Lettre de Monſieur Paſcal à M. de Fermat.

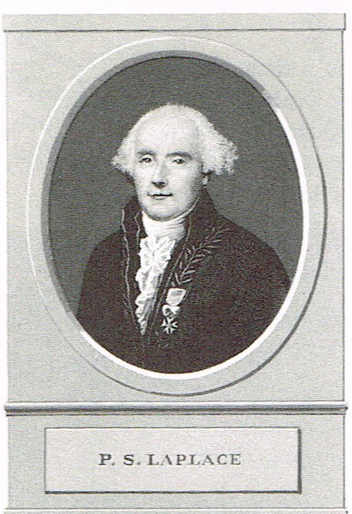
Du 24. Aouſt 1654.

MONSIEUR,

Je ne peus vous ouvrir ma penſée entiere touchant les partis de pluſieurs joüeurs par l'Ordinaire paſſé, & mémes j'ay quelque repugnance à le faire, de peur qu'en cecy cette admirable convenance qui étoit entre nous,& qui m'étoit ſi chere ne commence à ſe dementir,car je crains que nous ne ſoyons de differens avis ſur ce ſujet. Je vous veus ouvrir toutes mes raiſons, & vous me ferez la grace de me redreſſer ſi j'erre, ou de m'affermir ſi j'ay bien rencontré. Je vous le demande tout de bon & ſincerement, car je ne me tiendray pour certain que quand vous ſerés de mon côté.

Début de la lettre de Pascal à Fermat sur un problème de probabilité en matière de jeux.

Prenons le joueur de dés, exemple traditionnel auquel il nous faut bien nous référer. Pourquoi dit-on qu'il y a une chance sur six de sortir n'importe quel chiffre, de 1 à 6? *D'abord par expérience*, et cette remarque est fondamentale. La notion de probabilité est une notion expérimentale. Depuis les plus anciens temps, les joueurs ont constaté que le « un » ne sort ni plus ni moins souvent que le « deux », c'est-à-dire qu'en moyenne il sort une fois sur six. D'où résulta l'idée d'estimer la probabilité par la fréquence de sortie des divers événements possibles : c'est très exactement ce que permet la loi fondamentale des probabilités (**loi des grandes séries,** affirmant que les fréquences tendent presque certainement vers les probabilités). Puis des esprits pensants ont réfléchi aux raisons qui répartissent (et généralement équipartissent) les chances au jeu de dés et empêchent les coups d'être prévus. C'est que le coup joué est le produit d'un nombre immense de facteurs, bref le résultat d'une multitude de causes. Le mouvement du dé, sous l'action de la pesanteur et du frottement, obéit à la mécanique. Le chiffre obtenu dépend ainsi de la position du dé dans la main, de la forme de la paume, de son humidité, de la distance de la main au tapis, de la nature très complexe (une multitude de muscles interviennent) de l'impulsion imprimée au dé en le lançant. Après une chute dans l'air réglée par l'aérodynamique, l'impact fait intervenir l'élasticité du tapis, la dureté du dé, les irrégularités de toutes les surfaces en contact, en fait les lois des chocs et des frottements. Cette suite d'opérations a pourtant obéi à un déterminisme rigoureux; une intelligence supérieure avertie de tous les facteurs gagnerait à coup sûr.



C'est ce que le célèbre mécanicien français Laplace (1749-1827) a énoncé très nettement et très fortement :

« Une intelligence qui, pour un instant donné, connaîtrait toutes les forces dont la matière est animée et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'analyse, enchaînerait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle, et l'avenir comme le passé seraient présents à ses yeux. »

Remarquons la restriction : combien vaste devrait être cette intelligence; seul le recours à de puissantes et coûteuses machines électroniques permet de résoudre effectivement le problème du mouvement de trois ou quatre corps liés par leurs seules attractions mutuelles. Et il s'agit de résoudre un système d'autant d'équations et d'inconnues qu'il y a de particules dans l'univers, interagissant par bien d'autres forces que celles de gravitation! (Nous verrons ultérieurement une autre restriction essentielle : une telle analyse n'est possible que si l'état initial est connu exactement, ce qui est irréaliste, ainsi qu'il résulte évidemment de la nécessaire incertitude entachant toute mesure.)

Même si l'on peut, à bon droit et ainsi que le propose le plus élémentaire bon sens, négliger l'action de Vénus, ou de telle nébuleuse spirale immensément plus lointaine encore, sur le mouvement du dé, il est clair que les facteurs agissant certainement sur le résultat, tels qu'ils viennent d'être énumérés (sans même que l'on soit sûr de les avoir tous inventoriés), sont déjà beaucoup trop nombreux et complexes pour que l'analyse en soit possible. Surtout pendant le peu de temps dont nous disposons pour opter en faveur de pile ou de face.

Il résulte que, dans sa première caractéristique, le calcul des probabilités est l'outil qui permet d'analyser les effets résultant de causes multiples trop nombreuses pour être connues, étudiées et traitées séparément. C'est en cela qu'on pourrait le tenir pour un constat d'ignorance, ou plutôt d'incapacité, car l'homme n'ignore ni la nature des causes ni les lois selon lesquelles elles agissent; il manque seulement du temps nécessaire pour en combiner les effets.

Il n'est guère réaliste d'espérer prévoir le mouvement d'une molécule d'azote dans l'atmosphère : il y a bien trop d'autres particules, trop de chocs avec elles; et cependant les lois des rebondissements sont bien connues.

Personne, en outre, ne s'intéresse à telle ou telle molécule d'azote, mais bien au choc répété de milliards et de milliards d'entre elles sur une paroi, c'est-à-dire à la pression. On conçoit qu'il s'agisse là d'une « moyenne ». La loi des grandes séries prend en l'occurrence la forme des lois de Mariotte et de Gay-Lussac. Pareillement, l'autorité militaire ne porte aucune attention au fait que le conscrit Dupont, de la future classe 75, portant tels et tels gènes et chromosomes de ses ascendants, mesurera 1,76 m. Mais il lui plaît de savoir que 45 p. 100 des recrues de ce contingent seront plus grandes, fût-ce pour commander les tailles des uniformes.

Telle apparaissait la nature du calcul des probabilités à l'aube de ce siècle : une mise en ordre des phénomènes à causes multiples sans prépondérance trop notable de l'une d'elles (un dé dont une face serait considérablement alourdie ne donne plus lieu à probabilité, mais à certitude : celle d'être « roulé »). Il était d'ailleurs évident à beaucoup que, le progrès aidant, nombre de faits seulement probables deviendraient prévisibles à coup sûr. Quoi qu'il en soit, la seule hypothèse — dite « de Laplace » — de l'intelligence supérieure embrassant toutes les causes et prévoyant en conséquence tous les effets ôtait toute inquiétude aux théoriciens : **le monde était probabiliste par difficulté d'analyse, et non par principe.**

Il ne reste plus rien de ces belles espérances. Peut-être un peu trop métaphysiques, elles se sont évanouies, comme ont disparu les concepts de temps absolu indépendant de l'espace qu'il sert à mesurer, d'expérience indépendante des instruments de mesure, d'expérience indépendante de l'homme qui l'exécute. **Le monde est devenu probabiliste par essence.**

Qu'est-ce à dire?

Ceci, que l'hypothèse de Laplace, postulant la possibilité de connaître parfaitement l'état d'un phénomène à un instant et d'en déduire rigoureusement son évolution certaine, est contraire à la théorie et à l'expérience. Toute mesure, toute observation perturbe l'objet observé. Nous avons déjà vu en mécanique le principe de l'égalité de l'action et de la réaction; de même, ici, on ne peut être informé qu'en étant déformé par les informations reçues, c'est-à-dire finalement au prix d'une déformation équivalente de l'objet étudié.

Certes, on ne modifie guère la trajectoire d'un « Mystère IV » en le bombardant d'ondes radar ou de photons, tant sont différents les ordres de grandeur des êtres dont interagissent les quantités de mouvements. Mais que ferait un joueur de billard qui ne pourrait déterminer la position des boules qu'en les bombardant avec des objets de même importance? Gagnerait-il souvent s'il admettait que cette mesure ne modifiait rien et que la boule à atteindre restait là où elle était initialement?

Paradoxe, direz-vous; vit-on jamais joueur si inconscient et mesure si destructive. Remplacez « boules de billard » par « particules élémentaires » (ou « très petits objets »), reconnaissez qu'il n'existe pas de particules plus petites que les particules élémentaires, et relisez les dernières phrases du paragraphe précédent en portant attention aux mots « objets de même importance ». *Le mécanicien des nucléons ou des petits systèmes n'est pas dans la situation confortable du spécialiste de D. C. A.*

où l'on retrouve une vieille connaissance : la limitation de l'information

C'est ce que traduisent très précisément les relations d'Heisenberg : on ne peut connaître à la fois parfaitement les deux éléments nécessaires pour préciser le futur d'une particule, position et vitesse. En conséquence, l'avenir de celle-ci est fondamentalement incertain. Sans doute un monde inobservé pourrait-il poursuivre rigoureusement son évolution; mais, par définition, ce monde n'existerait pas pour nous.

Là-dessus, il ne convient nullement d'entrer dans une querelle où le raisonnement le cède à la passion. Cette incertitude d'évolution fut bien sûr une brèche dans le déterminisme laplacien, et exploitée comme telle à des fins diverses. Il ne semble pourtant pas qu'une telle incertitude traduise une ignorance passagère, mais bien une propriété essentielle. Pourquoi s'en étonner? On ne peut, avec un double-mètre, mesurer qu'avec une grosse incertitude la position des objets. De même, quel que soit l'instrument considéré, eût-il l'extrême finesse de la lumière, la précision sera nécessairement limitée. Nous n'avons aucun motif scientifique de faire des conjectures à propos d'un monde observé avec des instruments de dimension nulle. N'ayant pas d'effet sur lui, ils ne nous apprendraient rien.

Aussi l'histoire des probabilités est-elle très simple : à leur naissance, un aimable mais savant divertissement pour mathématiciens du XVIIe siècle raisonnant sur la théorie de la roulette et des jeux (dès lors, probabilistes et statisticiens ont extrait de la notion de jeux celle de risques à courir nécessairement, d'autant plus dangereux que l'on est ignorant : s'informer permet de jouer à coup plus sûr; **la statistique est la science des risques et des décisions**); jusqu'aux années 1900, un outil puissant d'appréhension des phénomènes à causes très multiples; depuis, l'expression même du fait scientifique élémentaire, et, partant, l'unique formulation possible des lois; le tout exprimé dans le langage le moins bavard et le plus universellement compris, celui des mathématiques.

l'exploitation des grandes séries

Il en va du calcul des probabilités comme de l'astronomie. Avant de parvenir à des résultats pratiques et souvent fort attrayants, il faut absorber un certain volume de matériaux relativement austères. L'astronautique suppose la cosmographie, et les probabilités, une certaine connaissance de la science du dénombrement appelée « analyse combinatoire ».

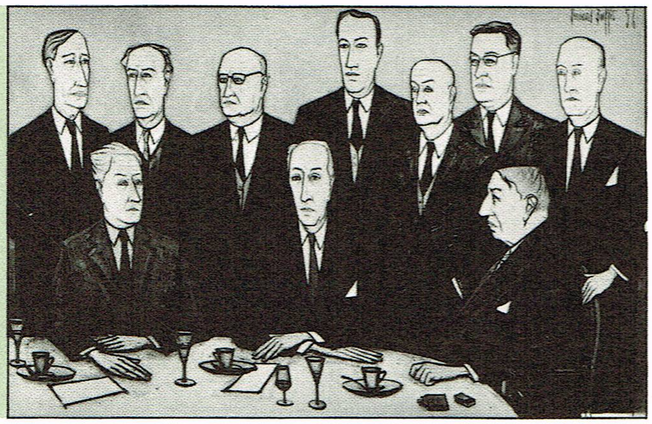
problèmes de trains et de préséance

Nous en faisons d'ailleurs tous les jours, et notamment en plaçant à table nos invités. Les Goncourt se réunissent à dix et, s'ils n'obéissaient à des règles de préséance sans doute assez rigides, il va de soi que le président pourrait occuper une des dix places, le plus âgé une des neuf places restantes, le suivant une des huit, etc. Ils ont ainsi la faculté d'occuper la table chez Drouant de $10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ façons différentes (soit 3 628 800). Par commodité d'écriture, le nombre $10 \times 9 \times 8 \times$... s'appelle « factorielle 10 » et s'écrit « $10!$ ».

Le jury du Goncourt peut siéger de 10! manières, comme un train à 10 wagons peut être composé de 10! façons différentes. On dit qu'on a affaire à des permutations; l'ordre et la nature des objets concernés interviennent, et il n'est pas indifférent que la voiture-restaurant soit en tête ou au milieu du convoi.

En revanche, si le train comporte 3 voitures de première et 7 de deuxième classe, il est généralement indifférent au voyageur qui n'a pas loué sa place que ce soit telle ou telle voiture de première qui soit en tête, pourvu qu'il en ait une. La S. N. C. F. peut alors atteler une première voiture de première classe à l'une des dix places après la motrice, la deuxième à l'une des neuf places

Lorsque les membres de l'académie Goncourt (ici, vus par Buffet), se réunissent au restaurant Drouant ils auraient $10! = 3\,628\,800$ façons différentes de se placer à table sans cérémonie.



Une fonction croissant vertigineusement : *n!*

1! =	1	6! =	720	20! ~ 2,1	10^{18}
2! =	2	7! =	5 040	50! ~ 3,0	10^{65}
3! =	6	8! =	40 320	100! ~ 9,3	10^{158}
4! =	24	9! =	362 880	1 000! ~ 1,0	$10^{2\,568}$
5! =	120	10! =	3 628 800	1 000 000! ~ 8,2	$10^{5\,565\,709}$

il faut plus de 5 millions de chiffres pour écrire l'entier 1 000 000! Quatre tomes de l'importance de cet ouvrage n'y suffiraient pas.

restantes, la troisième à l'une des huit places, etc., soit $10 \times 9 \times 8$ cas possibles. Mais le voyageur, si les 3 voitures en question ont été rangées aux places n°s 3, 4, 9, n'est pas intéressé de savoir si la voiture n° 1 occupe la troisième place plutôt que la quatrième ou la neuvième.

Bref, il ne s'occupe pas de l'ordre dans lequel sont rangées entre elles les voitures de première. Cet ordre est une permutation de 3 objets, soit 3!

En définitive, pour ce voyageur, il n'y a que $\frac{10 \times 9 \times 8}{3!}$ trains possibles; chaque convoi est appelé « combinaison ». Chacune d'elles a consisté à placer 3 voitures à 10 places possibles, sans tenir compte de l'ordre. Symboliquement, on écrit

$$C_{10}^{3} = \frac{10 \times 9 \times 8}{3!} = \frac{10!}{3!\,(10-3)!} = 120.$$

Voilà pour l'analyse combinatoire, qui prête évidemment à de nombreuses variations. Les exemples qui suivront permettront de se familiariser avec elle.

qu'est-ce qu'une probabilité ?

Peu de notions sont aussi instructives, et d'apparence aussi simple, que celle de probabilité. Qui nierait que la probabilité de sortir pile au prochain jet d'une pièce de monnaie est $\frac{1}{2}$, du moins si personne ne triche, si la pièce n'est pas pipée et si elle est de bon aloi ?

En effet, il n'y a pas plus de raisons de voir sortir pile plutôt que face; les probabilités sont donc égales. De même, si rien ne permet de soupçonner la régularité d'un jeu de cartes, le valet de pique a autant de raisons de sortir que toute autre carte, soit exactement 1 chance sur 52. Quant à la probabilité de tirer un pique quelconque, et non plus nécessairement le valet, elle vaut bien évidemment C_{52}^{13}, puisqu'il y a treize cartes « favorables » différentes.

On voit déjà apparaître les règles fondamentales :

1° *Une probabilité est un nombre compris entre 0 et 1. Il est d'autant plus grand que l'événement est plus probable (au sens habituel de ce terme)* ;

2° *La probabilité que se produise un parmi plusieurs événements incompatibles entre eux (le tirage d'un valet de pique exclut celui d'un roi de pique, ou de tout autre pique, puisque l'on ne choisit qu'une carte) est la somme des probabilités de chacun* (1).

Il est alors possible de construire tout le calcul des probabilités à partir de ces seuls axiomes; comme toute construction abstraite et axiomatique, cela comporte des risques, si l'on oublie par trop le sens physique de la théorie.

Nous obtenons finalement une définition de la probabilité d'un événement : c'est

---

(1) On dit plus mathématiquement que la probabilité est une fonction positive et additive, comme l'est la masse en mécanique (la masse d'une somme est la somme des masses partielles). Et, en fait, de fructueux rapprochements sont possibles entre mécanique et probabilité, et ce en dépit des différences de points de vue (une moyenne et une dispersion se définissent comme un centre de gravité ou un moment d'inertie).

Pour le voyageur n'ayant pas loué de place, ces deux convois de dix voitures ne présentent aucune différence. L'ordre entre elles des voitures de 1re et de 2e classe n'importe pas : il y a combinaison et non permutation. Le voyageur ayant loué sa place s'intéresse aux permutations et les deux convois ne sont pas identiques. *Ci-dessous*, une permutation et une combinaison différente des voitures.



Phot. Giraudon.

Sept est six fois plus probable que deux.

Le dé est conçu pour que le centre de gravité soit au centre du cube (ce qui n'est possible que si l'on grave correctement les numéros et ce qui impose la position respective des faces). Les six faces sont équiprobables :

$$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6.$$

Si l'on jette deux fois le dé et si l'on totalise les points, 7 est plus probable que 2. Seule la combinaison 1 + 1 peut donner 2, alors que 7 peut provenir de six lancers différents :

1 + 6, 2 + 5, 3 + 4, 4 + 3, 5 + 2, 6 + 1.

Comme il y a 6×6 lancers possibles :

$$p_2 = \frac{1}{36} \qquad p_7 = \frac{6}{36}$$

et de même :

$$p_2 = \frac{1}{36} \quad p_3 = \frac{2}{36} \quad p_4 = \frac{3}{36} \quad p_5 = \frac{4}{36} \quad p_6 = \frac{5}{36}$$

$$p_{12} = \frac{1}{36} \quad p_{11} = \frac{2}{36} \quad p_{10} = \frac{3}{36} \quad p_9 = \frac{4}{36} \quad p_8 = \frac{5}{36}.$$



six manières d'obtenir 7

seule manière d'obtenir 2

le quotient du nombre de cas où l'événement se réalise au nombre de tous les cas possibles, chacun ne semblant pas avoir plus de raisons que l'autre de se réaliser. Ainsi, tirer le roi de cœur ne peut résulter que d'un seul cas favorable, celui où on le tire effectivement ($p = 1/52$), les autres cas étant ceux où l'on tire une carte quelconque. Le tirage d'un as a 1 chance sur 13 de se réaliser, car 52 cas sont possibles et 4 favorables ($p = 4/52$).

D'où un troisième axiome, si évident qu'on a pu le proclamer tautologique :

3° *Lorsqu'on n'a aucune raison de croire que le mode opératoire adopté encourage un résultat plutôt qu'un autre, on admet que les probabilités sont égales (équipartition des probabilités élémentaires).*

une belle série

Peu de bridgeurs peuvent se vanter d'avoir eu 9 atouts dans la même main. Combien y a-t-il de mains différentes comprenant 9 cœurs (et par conséquent 4 cartes d'une autre couleur)? Il y a C_{13}^9 manières différentes d'avoir 9 cartes fixées à l'avance (cœurs) parmi 13. Quant aux quatre cartes restantes, il faut les choisir parmi 39, ce qui assure C_{39}^4 possibilités. Il y a donc finalement $C_{13}^9 . C_{39}^4$ cas favorables.

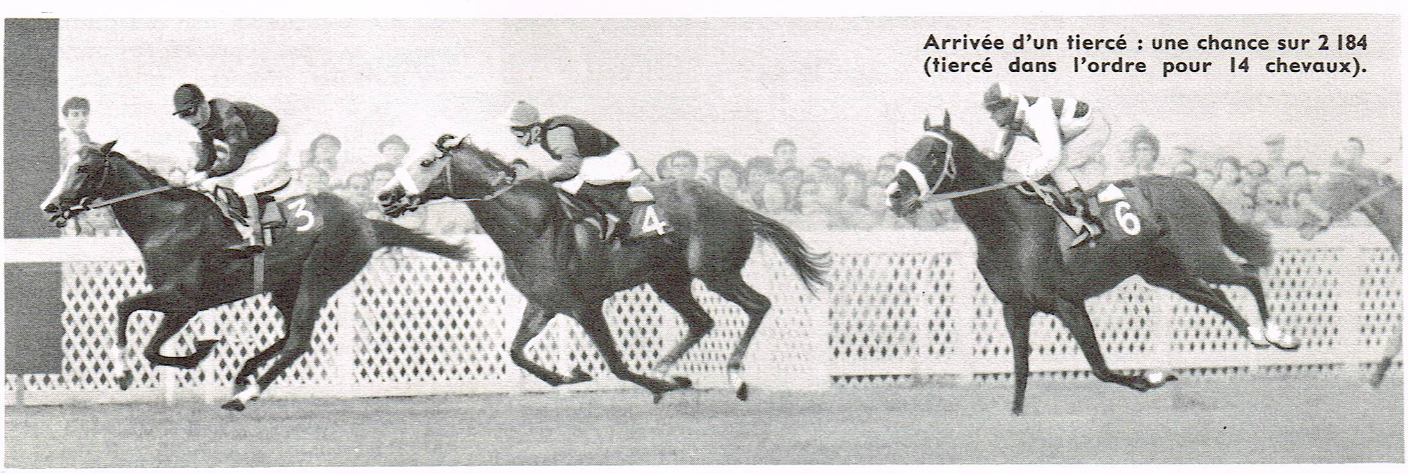
Et tout autant d'avoir 4 piques, ou 4 trèfles ou 4 carreaux. Il faut donc multiplier d'autant le résultat précédent, et l'on arrive à la probabilité $4 \, \frac{C_{13}^9 . C_{39}^4}{C_{52}^{13}}$ (soit un peu plus de 1 chance sur 3 000).

Précisons exactement le degré de rareté d'une telle donne : il faut jouer plusieurs milliers de fois pour espérer voir se produire en sa faveur un hasard si heureux.

le tiercé

Chacun connaît les paris « tiercés ». Certains se souviennent peut-être qu'à l'occasion d'un scandale intervenu à la fin de 1962 dans les milieux du turf parisien, certains journalistes en mal de copie avaient prétendu découvrir des « failles » dans la théorie de ce jeu. Il aurait peut-être été plus exact de parler de failles dans sa pratique, car, hélas! la théorie en est très simple : 14 chevaux, que l'on suppose d'abord d'égale valeur (équiprobables), prennent le départ. Quelle est la probabilité pour le ticket portant les numéros 3, 4, 6 dans l'ordre d'être gagnant? Le nombre de cas possibles est le nombre total de permutations de 14 chevaux, soit 14! Et le joueur va gagner lorsque les trois chevaux donnés arriveront premiers, accompagnés des 11 derniers, classés dans n'importe quel ordre. Il y a donc 11! cas favorables, soit une probabilité de

$$\frac{11!}{14!} = 12 \times 13 \times 14,$$



Arrivée d'un tiercé : une chance sur 2 184 (tiercé dans l'ordre pour 14 chevaux).

Phot. Keystone.

soit une chance sur 2 184. Si, bien sûr, le joueur n'avait pas été obligé de préciser l'ordre d'arrivée des 3 premiers, il eût augmenté le nombre des cas favorables, toute permutation des 3 premiers chevaux étant pour lui désormais gagnante. Il y en a $3! = 6$. Et le turfiste a 6 chances sur 2 184 ou 1 chance sur 364 de gagner.

Tout cela suppose que chaque cheval a la même chance d'arriver en tête. Mais chacun sait qu'il n'en est pas ainsi, car l'adresse du jockey, la valeur intrinsèque et la forme de sa monture, l'état du terrain font que Saxophone passera probablement Clysopompe, et il ne faut plus attribuer la même probabilité à chaque cheval. L'analyse des cas possibles et des cas favorables se complique un peu, mais sans changer ni de nature ni de principe.

Il faut tenir compte de l'agrément que ressent le joueur à la perspective de forcer le sort par un choix judicieux, en fonction des « tuyaux » qu'il possède, toujours de source sûre; et de l'espérance qu'il est ainsi en état de nourrir, jusqu'à la proclamation des résultats. Et, finalement, de ce qu'il ne touche que, divisée par le nombre de gagnants, la totalité des enjeux amputée des droits, des taxes et des frais — c'est-à-dire généralement assez peu. Peut-être faut-il insister sur ce point : il convient de bien distinguer entre les chances de gagner (qui posent un problème de pure probabilité et dépendent des paramètres propres de la course : cheval, jockey, terrain) et le gain effectivement encaissé (qui provient de la répartition des enjeux des autres parieurs). Cette répartition devrait suivre à peu près la cote que le public et les experts attribuent assez unanimement à un cheval, de sorte que le favori rapporte assez peu et l'outsider beaucoup. En fait, de nombreux turfistes dédaignent les gains faibles et jouent le « tocard » un peu plus souvent que ne le commanderait sa réelle valeur. D'où des surprises parfois amères, et cette constatation sans doute exacte qu'en jouant toujours les favoris on peut espérer un gain de 3 à 4 p. 100 par an.

Les probabilités examinées jusqu'ici ne concernaient que des événements simples. Il va de soi que des événements multiples — toucher trois fois de suite le tiercé, faire 421 puis tous les six, ou perdre 20 fois de suite à la loterie — rentrent aussi dans le champ des études probabilistes. Bien plus, c'est de la considération des événements répétés que naissent les grandes lois des probabilités, car l'outil mathématique, assez peu différent du sens commun dans les cas simples, prend toute sa puissance à débrouiller les situations inextricables.

Dans tous les ouvrages, et à cette place, on trouve alors le théorème des probabilités composées : si deux événements sont indépendants (c'est-à-dire que la réalisation du second ne dépend en rien de celle du premier), la probabilité qu'ils aient lieu tous les deux est égale au produit de la probabilité de chacun d'eux. Ainsi j'ai une chance sur quatre de tirer pile deux fois de suite, puisque cet événement résulte du concours de deux autres : pile la première fois (1 chance sur 2) et pile la seconde fois (1 chance sur 2). Il paraîtra normal à chacun d'admettre que la sortie de pile au deuxième jet de la pièce ne dépend pas de la première partie jouée.

Cependant, beaucoup hésiteront à jouer « rouge » à la roulette quand cette couleur est déjà sortie 15 fois de suite, ce qui arrive de temps en temps (1 fois en moyenne sur 32 768). L'attention et la réflexion du lecteur doivent être impérativement sollicitées à se porter là-dessus, car l'exemple qui vient d'être donné est le prototype des erreurs communes, et à vrai dire très humaines, qui entachent notre raisonnement.

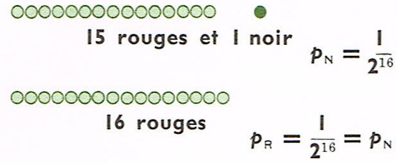
La roulette n'a véritablement, comme l'a dit Borel, ni conscience ni mémoire. *Peut-on prêter à une bille l'intention de modifier ses frottements, sa chute et son roulement pour s'arrêter de préférence sur un chiffre noir après avoir occupé 15 fois de suite une case rouge?* Et l'événement constitué par la sortie de 15 rouges suivis d'une noire a autant de chances de se produire que la sortie de 16 rouges successifs, chacune de ces probabilités valant 2^{-16}.

L'expérience, juge souverain, le prouve suffisamment et recoupe parfaitement les prévisions.

Il arrive fréquemment qu'une suite d'événements ne soient pas indépendants et que la réalisation de l'un commande en partie celle de l'autre. Ainsi cherchons quelles chances nous avons au bridge de retourner les quatre as de suite. L'événement composé est formé de la suite — ou produit — de quatre événements successifs qui consistent chacun à tirer un as. Le premier a la probabilité $p_1 = \frac{4}{52}$, mais non le second, puisque l'on doit supposer avoir déjà tiré un as. Il en reste 3 et 51 cartes, d'où $p_2 = \frac{3}{51}$. La probabilité cherchée est finalement de $\frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49}$ (1 chance sur 270 000). Voilà explicité le théorème des probabilités composées, portant cette fois sur des événements non indépendants : on fait toujours le produit des probabilités partielles, mais chacune d'elles tient compte de ce que l'événement précédent s'est produit. C'est ce qu'on nomme tout naturellement une « probabilité conditionnelle ».



Deux séries également probables : la boule n'a pas de mémoire.



Phot. Weiss, Rapho.

raisonnements sur des pointes d'aiguille

Reprenons le cas simple, déjà examiné à plusieurs reprises, du tirage à pile ou face d'une pièce loyale. Nous avons vu que, après 2 tirages, pile pouvait sortir 0, 1 ou 2 fois. Les cas extrêmes de tirage, tous favorables ou tous défavorables, peuvent donc se produire. Mais une certaine compensation semble se réaliser, puisque le tirage mixte (1 fois pile et 1 fois face) est deux fois plus probable que chacun des autres :

$$p_0 = \frac{1}{4}, \qquad p_1 = \frac{2}{4}, \qquad p_2 = \frac{1}{4}.$$

Ce résultat provient non de ce que les quatre tirages possibles ne sont pas équiprobables, mais de ce que plusieurs événements distincts sont favorables à la réalisation d'un seul succès (pile d'abord, puis face-face, puis pile).

De même, si l'on effectue trois tirages, on trouve :

$$p\ (0\ \text{face}) = \frac{1}{8}, \quad p\ (1\ \text{face}) = \frac{3}{8},$$

$$p\ (2\ \text{faces}) = \frac{3}{8}, \quad p\ (3\ \text{faces}) = \frac{1}{8}$$

(trois tirages sont favorables au deuxième cas, la « face » pouvant être tirée en premier, en deuxième ou en troisième lieu, alors que seul le tirage face-face-face est favorable au premier résultat);
et pour quatre :

$$p\ (0) = \frac{1}{16}, \quad p\ (1) = \frac{4}{16}, \quad p\ (2) = \frac{6}{16},$$

$$p\ (3) = \frac{4}{16}, \quad p\ (4) = \frac{1}{16}$$

(on vérifie que la somme des probabilités donne toujours l'unité).

On peut continuer indéfiniment l'analyse mathématique. Indiquons simplement le résultat :

$$\textbf{probabilité}\ (i\ \textbf{succès sur}\ n) = \frac{C_n^i}{2^n},$$

puisque le nombre total de cas est 2^n et qu'il y a autant de résultats favorables qu'il y a de manières de caser exactement i gains en n parties, c'est-à-dire C_n^i (l'analogie est complète avec la répartition de trois voitures de première classe sur une rame de dix voitures).

On peut aisément calculer les C_n^i (triangle de Pascal) et en déduire les courbes donnant $p\ (i/n)$: on remarque combien les valeurs voisines de $n/2$ deviennent rapidement quasi certaines.

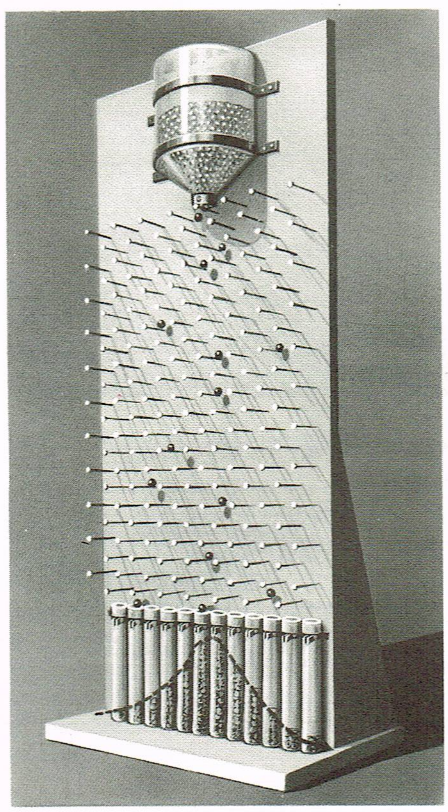
Un appareil simple, connu dans les foires et les kermesses sous le nom de « jeu des pointes d'aiguille », permet de matérialiser très facilement un tel résultat. La boule jetée à la verticale de la pointe O rebondit; A_1 et A_2 sont suffisamment larges et bien placés pour que la bille rebondisse une deuxième fois sur A_1 ou sur A_2, puis elle tombera sur B_1, B_2 ou B_3, et ainsi de suite jusqu'à sa chute dans une des cases finales. On demande de prévoir à l'avance le lieu de la chute. Et la réponse est facile : toutes les cases sont possibles, mais il vaut 6 fois mieux parier sur D_3 (auquel 6 chemins aboutissent) que sur D_1 ou D_5.

C'est ce que vérifie l'expérience. Si l'on procède à un grand nombre de jets, la case centrale est beaucoup plus remplie que les cases extrêmes, et ce d'autant plus que les chocs sont multiples :

Une certaine compensation s'opère déjà.

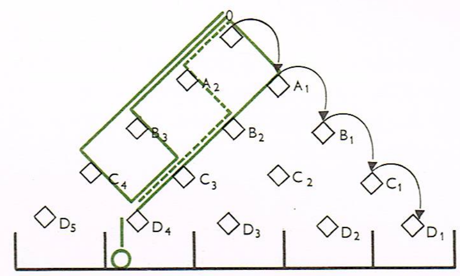


0 face **un seul événement est favorable** $p_0 = 1/4$	**1 face** **deux possibilités favorables** $p_1 = 2\ 1/4 = 1/2$	**2 faces** **un seul événement favorable** $p_2 = 1/4$

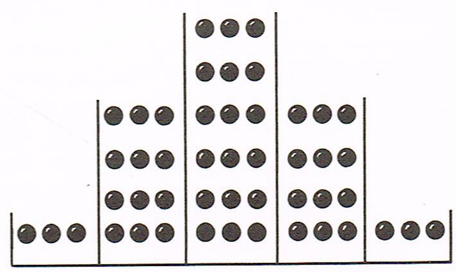


Le jeu des aiguilles en quinconce.

Une certaine compensation semble se réaliser : les valeurs voisines de $\frac{n}{2}$ (ici 1) deviennent rapidement quasi certaines.



Un seul trajet (noir) aboutit à la case extrême droite et quatre (verts) à la deuxième case.



État final moyen après 48 lancers.

Tableau de Pascal.

	0	1	2	3	4	5	6	7	
1	1	1							1 + 1 = 2
2	1	2	1						4
3	1	3	3	1					8
4	1	4	6	4	1				16
5	1	5	10	10	5	1			32
6	1	6	15	20	15	6	1		64
7	1	7	21	35	35	21	7	1	128

On observe le mode de formation du tableau (dû à Pascal), 2 + 1 = 3   5 + 10 = 15   20 + 15 = 35, la symétrie et la grande concentration autour de l'axe de symétrie.

6 fois plus remplie, s'il y a 4 rangées;
35 fois plus pour 7 rangées;
252 fois plus pour 10 rangées;
184 756 fois plus pour 20 rangées;
1 300 000 000 000 000 fois plus (environ) pour 50 rangées.

Pratiquement, donc, la concentration devient extrême quand *n* est grand, et *l'on est sûr de perdre en pariant pour une case extrémale.*

la loi des grandes séries, clé de l'architecture probabiliste

Enonçons un peu différemment le résultat précédent en considérant le numéro de la pointe heurtée : une boule issue de C_3 ne peut heurter que D_3 ou D_4; ce qui revient à dire qu'il faut compter le nombre de fois où l'indice augmente de 1 unité pour avoir la case finale; ou encore qu'il faut énumérer les tirages favorables à l'avancement de l'indice.

Ainsi on retrouve le jeu de pile ou face qui nous a servi de point de départ, et l'on a démontré le théorème fondamental de la science des probabilités :

Pour peu qu'un phénomène se répète un nombre suffisant de fois, on peut parier avec certitude que sa fréquence d'apparition sera voisine de sa probabilité. En vertu de cette loi **(loi des grandes séries),** « face », qui a une probabilité $\frac{1}{2}$, sortira pratiquement 1 fois sur 2 si l'on fait des millions de tirages — et de même le 13 de la roulette 1 fois sur 37 (il y a 37 cases, y compris le zéro).

On ne saurait assez insister sur cette loi : tout l'édifice grandiose et universel des probabilités est fondé sur elle. Sans elle, aucune stabilité n'est possible; aucune expérience ne pourrait être acquise. C'est elle seule qui nous permet d'estimer les probabilités (par les fréquences de sortie); sans elle, nous n'aurions que du désordre.

Que l'on nous comprenne bien. Tirons un million de fois une pièce de monnaie. Tous les résultats sont possibles; on peut même n'obtenir que des piles ou que des faces (la probabilité est exactement de $2^{-1\,000\,000}$, ce qui s'exprime par zéro suivi de plus de 300 000 zéros). Mais tous ne le sont pas également. Si l'on pouvait tenter l'expérience un très grand nombre de fois, on pourrait trouver par exemple 501 012 faces, ou bien 498 861; mais presque toujours les fréquences resteront extrêmement voisines de 0,5. On dit qu'elles sont très peu « dispersées » autour de la probabilité, ce que traduit le tableau suivant :

avec la probabilité	la fréquence de sortie sera comprise entre (1 000 000 de parties)		(10 000 parties)	
1 (à coup sûr)	entre 0 et 1		entre 0 et 1	
0,999 9	0,497 8	0,502 2	0,478	0,522
0,999 (99 fois sur 100)	0,498 5	0,501 5	0,485	0,515
0,99	0,498 7	0,501 3	0,487	0,513
0,90 (90 fois sur 100)	0,499 2	0,500 8	0,492	0,508
0,70	0,499 5	0,500 5	0,495	0,505
0,50 (1 fois sur 2)	0,499 7	0,500 3	0,497	0,503

ou encore le tableau suivant, équivalent :

la fréquence sera comprise		avec la probabilité (1 000 000 de parties jouées)	(10 000 parties)
entre 0 et 1		$p = 1$ (toujours)	$p = 1$ (toujours)
0,49	0,51	0,999 9	0,955
0,499	0,501	0,955	0,16
0,499 9	0,500 1	0,16 (16 fois sur 100)	0,01 (1 fois sur 100)

les politiques du risque calculé

Ainsi donc, aucun résultat n'est certain (sauf celui, évident et vide d'intérêt, que la fréquence est toujours comprise entre 0 et 1). Toute affirmation est risquée. On a simplement le choix entre une affirmation presque certaine, mais trop peu précise pour nous apporter une certitude enrichissante, et une affirmation très précise et intéressante, mais douteuse.

Nous atteignons là une des plus essentielles — pour ne pas dire l'essentielle — limitations de notre connaissance. *On ne peut vivre qu'en pariant, qu'en risquant. Et le bénéfice (en cas de gain) est à la mesure du risque accepté.*

Nous avons déjà vu beaucoup de bornes posées à nos possibilités de connaissance (toute information est nécessairement finie; toute connaissance est aléatoire; la capacité de transmettre des informations par un canal physique donné est également limitée), mais remarquons que l'expérience la plus courante nous impose quotidiennement l'obligation d'un risque perpétuel : toute action, que ce soit celle de traverser la rue, ou celle de quitter son lit, ou celle d'y rester, est risquée, et tout résultat est incertain. Tout au plus faut-il prendre des risques efficaces et fructueux. N'est-ce pas là le principal attrait de la vie?

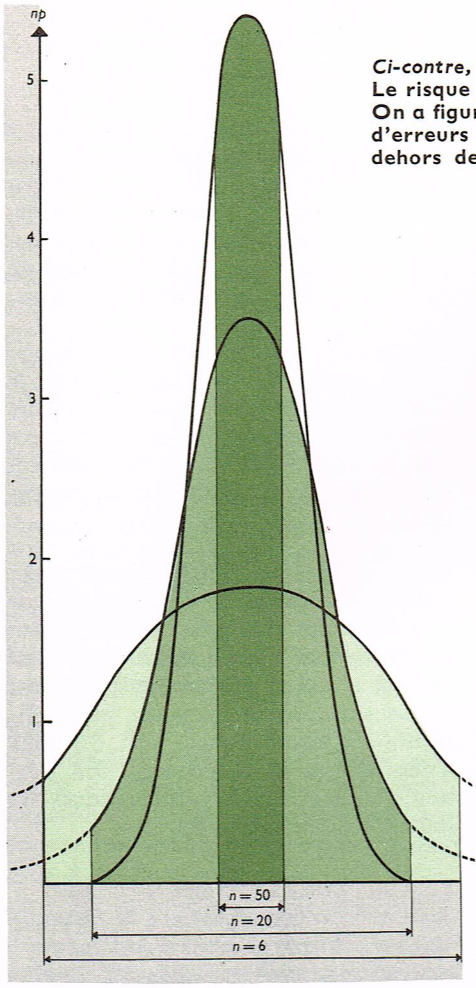
multiplier les expériences?

Mais, me direz-vous, il y a une solution permettant de rendre l'intervalle des fréquences aussi petit qu'on le souhaite, et la probabilité de succès aussi voisine de l'unité qu'on peut le désirer, le tout simultanément et sans avoir à opter : il n'y a qu'à multiplier les expériences. Faisons-en un million, cent millions même si ce n'est pas assez.

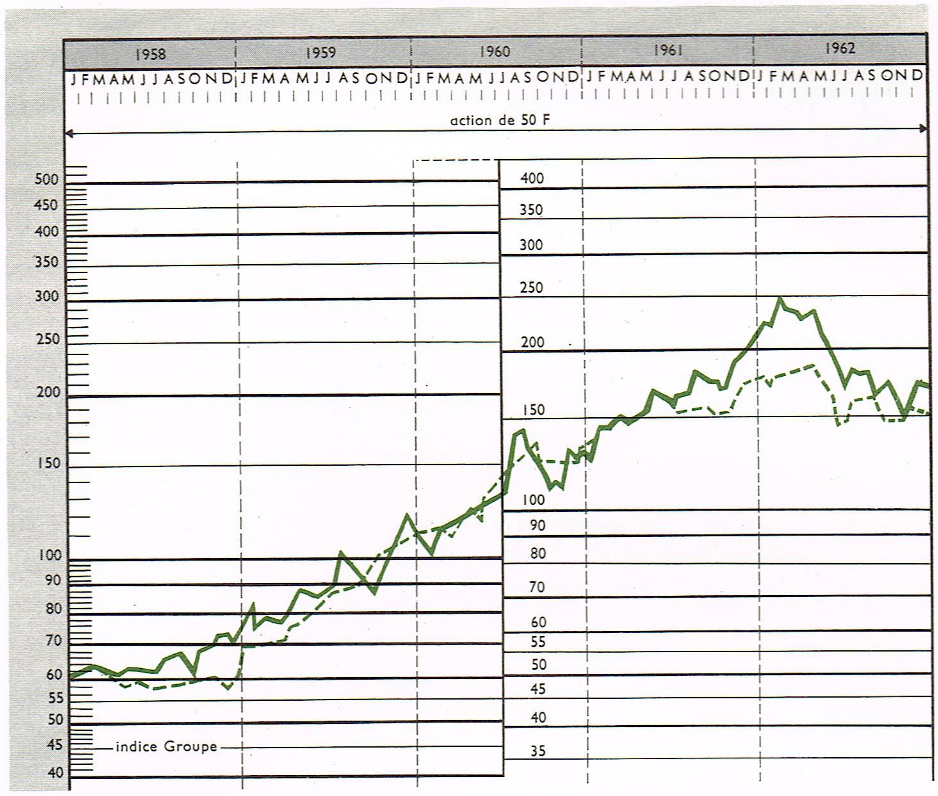
Il n'y a qu'à!...

N'insistons pas sur une première objection, tant elle est évidente et contraignante : multiplier les essais coûte cher et demande du temps! **Et le coût (en argent et en temps) intervient, quand il faut évaluer nos possibilités d'action; ici encore, nous ne sommes jamais ni infiniment riches ni immortels.**

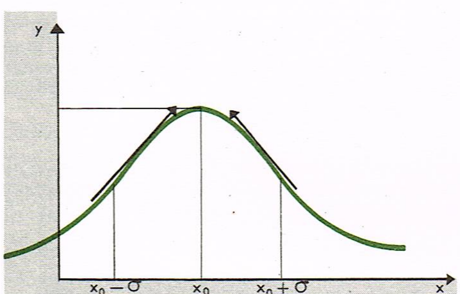
Une seconde difficulté, plus cachée peut-être, mais plus profonde, doit aussi nous retenir : lancer des millions de fois une pièce, plus généralement accumuler les expériences sur le même objet, amène une modification qui rend les observations non comparables entre elles, et le résultat dénué de sens physique. Si l'on admet que toute observation déforme la chaîne d'observation (observateur — instrument observant — observé), quelles propriétés observe-t-on : celles de la pièce neuve ou celles de la pièce usée par des millions de lancers?



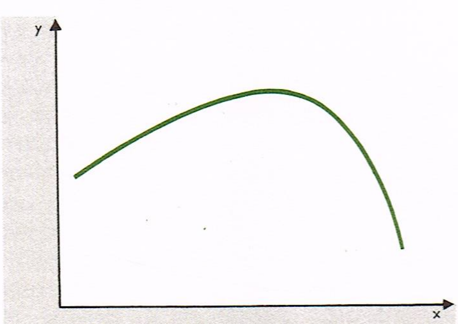
Ci-contre, courbes *np* permettant d'obtenir une fréquence de sortie au jeu de pile ou face. Le risque d'erreur (incertitude) décroît avec n : la courbe devient de plus en plus effilée. On a figuré — pour $n = 6$, $n = 20$ et $n = 50$ — la région centrale correspondant à 5 p. 100 d'erreurs possibles (on se trompe une fois sur vingt en pariant que la fréquence sera en dehors des zones hachurées). On remarque également que les courbes ont des formes voisines.



Un effet de moyenne : *ci-dessus* — en trait gras — le cours d'une action du groupe « matériel électrique » et l'indice moyen du groupe. On remarque le « parallélisme » des deux courbes (à part une petite pointe vers mars 1962 correspondant à une grosse commande passée à la société). On remarque surtout que l'indice a une marche nettement plus régulière : il s'agit d'un effet de moyenne.

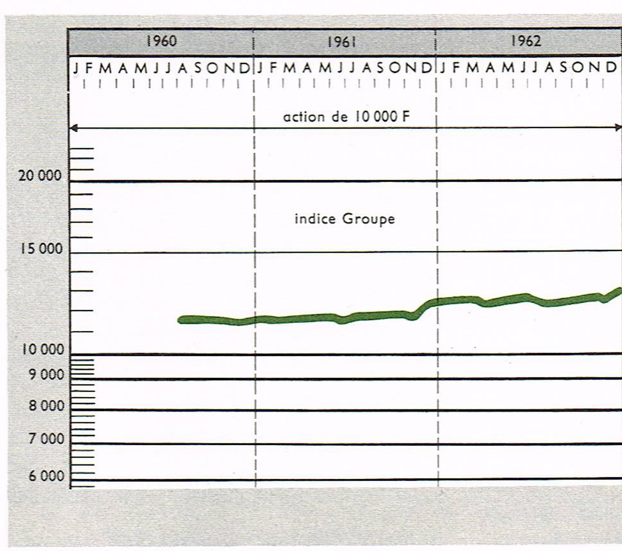


La variable normale la plus générale (de moyenne x_0 et de dispersion σ).



Une variable non normale

$\int f(x)\,dx = 1 \qquad f(x) \geqslant 0.$



Une société de portefeuille : le cours est beaucoup plus régulier encore, l'effet de moyenne jouant beaucoup plus fortement (cette société gère un portefeuille portant sur des actions très diverses. Et son cours reflète pratiquement le cours moyen de la Bourse.

Beaucoup complètent la loi des grands nombres par cet énoncé que « les événements très peu probables ne se produisent jamais ». Ce théorème — ou plutôt ce principe — a un petit côté « finaliste » assez dangereux pour l'esprit scientifique. Rien n'empêche les événements très peu probables de se réaliser, sinon l'extrême petitesse des chances qu'ils ont de s'accomplir. On cite volontiers le miracle, dit des « singes dactylographes » (pourquoi donc des singes?), qui, frappant au hasard sur le clavier d'une machine, reproduisent un vers ou un acte ou une pièce de Corneille. Quelles chances ont-ils d'y arriver? Si l'on admet 40 lettres en moyenne par vers, une sur 26^{40} d'écrire un vers donné (1 sur 10^{56}). Bien plus qu'une longue dissertation, ces chiffres astronomiques nous font sentir l'impossibilité d'un tel miracle. Il n'est pas facile d'écrire « le Cid » en employant un tel procédé.

les lois du hasard : de la « courbe en cloche » à la loi de Poisson

Regardons une dernière fois la loi de probabilité des gains au jeu de pile ou face, c'est-à-dire la loi donnant la probabilité de chaque événement possible représenté par un point de l'axe O*x*. Le nombre de gains est un nombre entier variant entre 0 et *n*, et, pour rendre plus lisibles les résultats, joignons-les par une courbe régulière, sans nous soucier des échelles. Nous voyons que de telles courbes ont rapidement (et, pratiquement, dès $n = 10$) des formes voisines. L'analyse mathématique le confirme, et précise la forme exacte de la courbe limite.

Avec des échelles convenables sur les axes des coordonnées, on trouve la loi limite

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

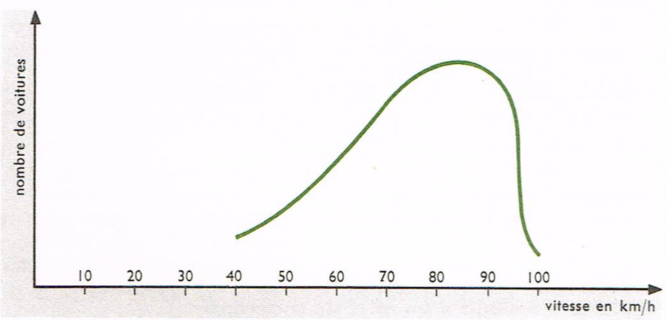
La découverte de cette loi a été attribuée à de Moivre (1667-1754), à Laplace (1749-1827) et à Gauss (1777-1855), trois grands savants qui l'ont progressivement révélée et exploitée. Aussi porte-t-elle le triple nom de loi de Laplace-Moivre-Gauss, ou, ce qui est plus simple et plus court à énoncer, celui de « loi normale » (la courbe la représentant étant appelée, plus vulgairement, mais de manière imagée, courbe « en cloche »).

Est-ce à la simple paresse qu'il faille attribuer l'octroi d'un nom aussi flatteur que celui de « loi normale »?

Eh bien, il se trouve que non, et que cette loi est la plus importante, la plus universelle (et, en dépit des apparences, la plus simple) de toutes les lois statistiques. *Elle caractérise tous les événements dont les causes sont multiples, indépendantes et à peu près de même importance.*

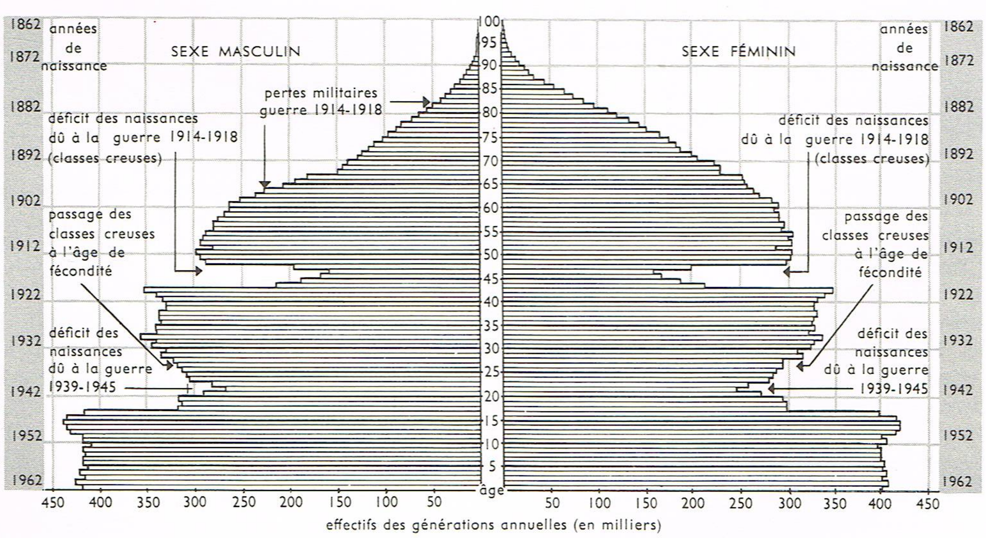
Notamment, les résultats d'une mesure — ou plutôt d'une longue série de mesures — sont généralement distribués selon une loi normale (les causes d'erreur sont en général multiples, mais souvent indépendantes et d'importance comparable).

Cette loi est si simple à manipuler mathématiquement et si universelle que l'on a commis des excès, la proclamant à tort valable alors que les clauses de validité n'étaient pas vérifiées, soit que les causes d'erreur n'étaient pas suffisamment nombreuses, soit qu'elles n'étaient pas indépendantes, soit surtout qu'une était prépondérante et imposait de ce fait sa loi.



La courbe donnant le nombre de voitures circulant sur une autoroute en fonction de leur vitesse s'effondre brusquement pour la valeur (90 km/h) correspondant à la limitation de vitesse imposée ce jour : la simple vue de la courbe permet de pressentir fortement l'existence d'un tel règlement, et de C. R. S. le faisant respecter à qui l'ignorerait *a priori*.

De même, *ci-dessous*, la pyramide des âges des Français traduit les effroyables hécatombes de la guerre (1914-1918 et 1939-1945) et la chute corrélative de la natalité.



estimer, et agir

un homme averti en vaut deux

Nous n'avons fait qu'entrevoir comment le calcul des probabilités pouvait déboucher sur la vie et l'action. Comment une discipline si mathématique et si abstraite, d'apparence essentiellement arithmétique, peut-elle permettre des estimations, des jugements, des choix, des actions? C'est maintenant l'heure d'appliquer le calcul des probabilités, c'est-à-dire de faire de la statistique.

la pièce était-elle loyale?

Et d'abord de nous poser deux questions simples. Comment calculer les probabilités? Que veut exactement dire l'affirmation « Face a une chance sur deux d'apparaître »?

Nous avons déjà donné deux réponses à ces questions naïves :

Réponse A : il n'y a aucune raison d'avantager face plutôt que pile, ou pile plutôt que face, si la pièce est loyale, symétrique. Donc pile et face ont des probabilités égales, donc égales toutes deux à 0,5;

Réponse B : la fréquence tendant vers la probabilité, il suffit de faire un nombre suffisant d'expériences pour pouvoir estimer en toute sécurité la probabilité par la fréquence. Cette dernière étant voisine de 0,5, on est amené à imposer la probabilité 0,5 à l'événement « face ».

Deux réponses? C'est trop, à moins que toutes deux ne traduisent la même réalité. Ou à moins qu'elles ne soient de fausses, ou du moins des pseudo-réponses. Il n'est donc pas inutile de les approfondir un peu.

En toute rigueur, la première ne peut qu'être approximative. Pile ou face ne sont pas rigoureusement identiques. Sinon, rien ne permettrait de les reconnaître et l'on ne saurait jamais laquelle est pile, et laquelle est face. *C'est justement parce qu'elles ne sont pas identiques physiquement que l'on peut les reconnaître.* Et alors l'hypothèse d'équiprobabilité devient suspecte : la différence ne peut-elle favoriser un côté si peu soit-il ? (Effectivement, comme l'a montré Weldom, c'est bien ce qui s'était produit avec les premiers dés : les points numéros étant creusés dans la matière du dé, la face du six était la plus légère. Elle apparaissait donc plus souvent que celle du 1, plus lourde, qui avait tendance à retomber plus souvent contre la table.) La réponse A ne peut être qu'approchée, et provisoire.

Il en est de même pour la seconde réponse : la fréquence tend presque certainement vers la probabilité, mais ne lui est pas nécessairement égale. En retournant les conclusions d'un paragraphe précédent, on peut seulement dire, à la suite d'un nombre nécessairement fini d'essais : il est presque certain (à 99 p. 100) que la probabilité est comprise entre 0,49 et 0,51. Sans, pour cela, espérer jamais une réponse exacte.

Est-ce à dire qu'il n'y a aucun espoir?

Il n'y a certes aucun espoir de connaître rigoureusement une seule probabilité exacte. Combien de fois n'avons-nous pas vu que toute connaissance est nécessairement limitée et aléatoire, et toujours sujette à révision, ou, mieux, à perfectionnement?

que faire alors ?

Mais le calcul des probabilités n'est pas un pur et gratuit jeu de l'esprit. Il est fructueux, si on sait l'utiliser habilement et sûrement.

Dans l'exemple cité, on commence par vérifier que la pièce de monnaie est aussi loyale que possible. (Si elle encourage un événement, elle ne doit donc pas l'encourager de manière sensible.) Admettant que $p = 0{,}5$, on en déduit toutes les conclusions utiles, par exemple qu'au bout de 1 000 000 de tirages indépendants « face » sortira entre 499 000 et 501 000 fois au moins dans 95 p. 100 des cas. On fait le tirage. Si face est sortie 499 327 fois, on observe un accord suffisant entre la fréquence effective et 0,5 pour que l'on puisse continuer à faire confiance à la loyauté de la pièce. Si face est sortie 501 507 fois (ce qui n'est possible que 1 fois sur 20 pour une pièce loyale), on peut commencer à soupçonner une irrégularité — volontaire ou non.

Pourquoi avoir choisi le seuil de 95 p. 100, et non celui, plus exigeant, de 99 p. 100, par exemple? Pourquoi accepter un risque de 5 p. 100 et non de 1 p. 100?

Cela est une question de politique, et notre choix doit être guidé tant par notre esprit d'aventure et de risque que par tout ce que nous savons, ou soupçonnons, par ailleurs. Joue-t-on avec un voisin de café, on fera plus attention que si l'on a comme adversaire un ami intime qui nous a donné de nombreuses preuves de sa loyauté. Si le premier bénéficie successivement de trois tirages victorieux, on sera plus tenté d'arrêter le jeu que si c'est le second qui a eu la main si heureuse. De même, plus l'enjeu est fort, plus vite une série d'échecs nous rendra prudents : un risque de 5 p. 100 peut être raisonnable s'il s'agit de perdre quelques centaines de francs; un risque de 1 p. 100 est totalement déraisonnable si l'on joue sa vie ou un bien très cher — à moins que l'enjeu n'en vaille la peine.

les pièges de la sélection observationnelle, ou la difficulté de choisir au hasard

Encore ne peut-on parler de statistiques, ou de fréquence, qu'en les rapportant à un certain ensemble d'objets. C'est une bonne chose que d'estimer, une meilleure encore que de savoir ce qu'on estime; « quel est l'éclat moyen des étoiles? » est une question dépourvue de sens si l'on ne précise pas : « des étoiles géantes » ou « des étoiles naines ».

Il faut donc partir d'une population (dite « parente ») bien définie. Ou, à défaut, et faute de pouvoir faire un inventaire complet et coûteux, étudier certains éléments convenablement choisis (échantillons).

Même lorsque la population parente est formée d'un nombre fini d'éléments (l'ensemble des Français, par exemple), il peut être coûteux ou même impossible de la considérer en entier. On procède alors à des sondages.

les sondages d'opinion

Un chef d'Etat désireux de connaître l'opinion à la veille d'un référendum, d'un scrutin ou d'une décision politique importante, ne peut demander à ses préfets d'interroger chaque Français (ce qui reviendrait à faire un vote pour

connaître les résultats que donnerait sensiblement la même consultation quelques jours après); il fera procéder à un sondage d'opinion, suffisamment important pour que l'on puisse se fier à l'analyse du résultat (par application de la loi des grandes séries), suffisamment restreint pour que le coût n'en soit pas prohibitif; il consultera par exemple 500 Français.

Mais comment choisir l'échantillon des 500 personnes à interroger?

On pourrait par exemple ouvrir un annuaire de téléphone (liste alphabétique) à la page 253 et prendre note des 500 premiers noms inscrits.

Mais alors on aurait l'opinion moyenne de 500 Parisiens (si l'on a pris l'annuaire de Paris) suffisamment riches pour posséder le téléphone. De plus, la liste comprendrait un excès d'Alsaciens, de Vietnamiens et de pharmaciens (la page 253 correspond à « Pf », « Pham » et « Pharm »). On connaîtrait peut-être la pensée politique (ou plutôt le vote) de bourgeois, d'Alsaciens et de Vietnamiens, mais croire qu'un référendum national suivrait les mêmes règles risquerait d'être désastreux.

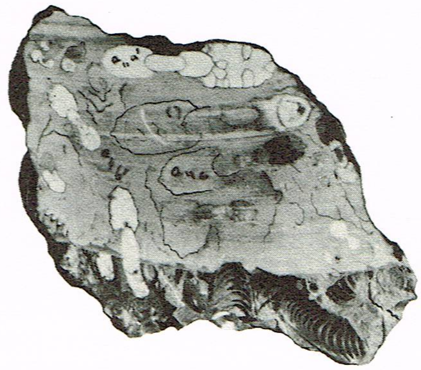
Le lecteur pourra examiner d'autres techniques de sondage. C'est un art difficile que d'aboutir à une population où sont équitablement représentés tous les âges, tous les milieux sociaux, toutes les professions, toutes les zones géographiques, tous les états de fortune, les deux sexes...

La critique est toujours facile; l'art est plus difficile et le bon sens doit nous guider. Un recours aux plus puissantes machines électroniques pour tirer des « nombres au hasard » est nécessaire aussi. Ces techniques de sondage ont été perfectionnées à la suite d'échecs particulièrement retentissants. Et l'on peut admettre aujourd'hui que, si un résultat de scrutin dément ce qui était prévu par les organismes les plus sérieux, il faut en accuser plus un changement d'opinion de dernière heure toujours possible du corps électoral qu'une insuffisance de la technique de sondage.

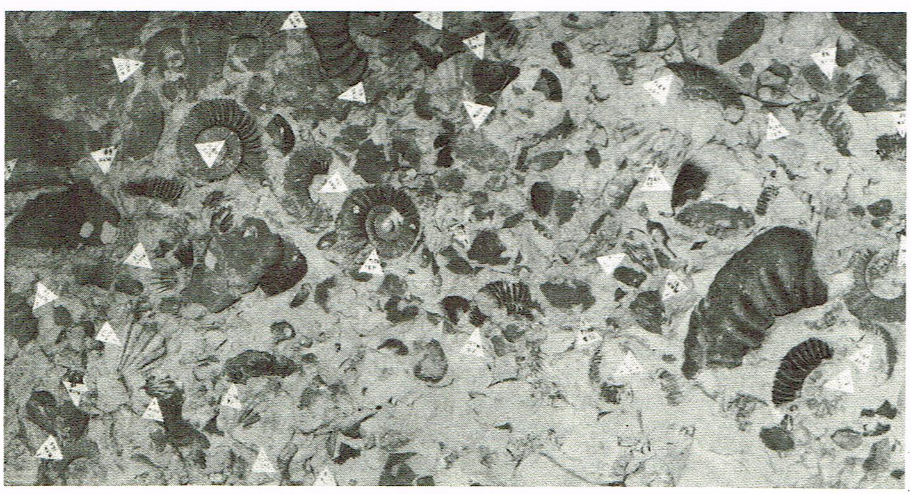
quelques problèmes de la préhistoire

Peu de sciences échappent à des problèmes analogues dus aux difficultés d'échantillonnage. Et surtout pas la paléontologie, qui étudie les êtres anciens, les organismes ayant vécu aux époques antérieures aux temps actuels. Les documents sont fournis par les restes, que l'on appelle les « fossiles ».

Une première difficulté résulte de leur rareté, sinon de leur absence. Ainsi, un heureux concours de circonstances a permis de connaître dans les terrains du

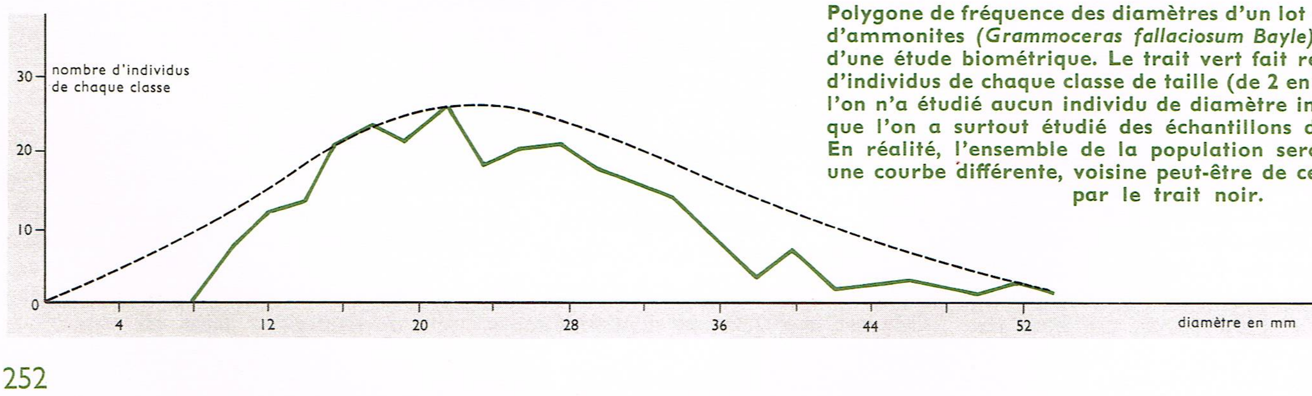


Section de fragment de nodule d'ammonites provenant de l'alésien des environs d'Obernai (Bas-Rhin). Dans ce nodule se trouvent réunis près d'une centaine d'échantillons d'ammonites appartenant à une seule espèce *(Pleydellia abensis Zieten)*. Noter qu'il y en a de toutes les tailles. Cet échantillon donne probablement une image assez fidèle de l'ensemble de la population.



Collovien moyen de Palente, près de Besançon.

Sur cet échantillon sont rassemblés un grand nombre d'ammonites et de lamellibranches. Leur disposition évoque l'idée d'une plage marine où les fossiles auraient été rassemblés par le jeu des vagues ou des courants. Il y a donc une sorte de tri par des agents physiques.



Polygone de fréquence des diamètres d'un lot de 285 échantillons d'ammonites *(Grammoceras fallaciosum Bayle)* ayant fait l'objet d'une étude biométrique. Le trait vert fait ressortir le nombre d'individus de chaque classe de taille (de 2 en 2 mm). Noter que l'on n'a étudié aucun individu de diamètre inférieur à 8 mm et que l'on a surtout étudié des échantillons de taille moyenne. En réalité, l'ensemble de la population serait distribué selon une courbe différente, voisine peut-être de celle qui est figurée par le trait noir.

cambrien, vieux de près de 500 millions d'années, des restes de méduses, organismes très fragiles. D'autres sont connus dans les terrains du trias, vieux de près de 180 millions d'années, et encore dans ceux du jurassique de Solenhofen, datant de 140 millions d'années. Dans l'intervalle, ce groupe a certainement existé dans de nombreuses mers, mais aucun reste ne nous est parvenu. Ainsi, après leur mort, la plupart des organismes disparaissent sans laisser de traces, leur matière organique étant rapidement détruite par les processus de la putréfaction.

Pourtant, certains gisements font exception. Les glaces de Sibérie ont fourni des centaines de rhinocéros laineux et de mammouths, ces derniers parfois en si bon état de conservation que l'on a même pu soumettre leurs tissus à des réactions sérologiques. Pareillement, l'ambre de la Baltique renferme des insectes dont on a pu faire l'étude histologique, reconnaître la structure des cellules.

Les constituants organiques sont donc en général détruits. Pourtant, certains d'entre eux sont pratiquement inaltérables. Tel est le cas de la chitine qui imprègne le corps des arthropodes, et notamment des insectes. Parmi les plus anciens fossiles bien conservés et utilisables par les géologues pour dater les terrains se trouvent précisément des arthropodes au corps divisé en trois parties : les trilobites. C'est leur carapace de chitine qui a permis la conservation de leurs restes, même de ceux qui se sont détachés au cours des mues que subissaient ces animaux.

Mais la plupart des organismes ne laissent que les parties dures : leur squelette calcaire comme celui des mollusques, leur test siliceux comme celui des éponges. Les dents couvertes d'émail se conservent facilement; ce sont les seuls restes que nous ayons de la plupart des vertébrés.

Par ailleurs, les meilleures chances de conservation sont offertes aux organismes vivant en milieu aquatique; à leur mort, ils tombent parfois au fond des eaux dans une vase azoïque (impropre à la vie), qui leur assure une fossilisation convenable.

En revanche, les animaux à vie aérienne ont peu de chances d'être conservés. Du calcaire lithographique de Solenhofen (jurassique) on n'a retiré que trois exemplaires d'oiseaux fossiles. De la période suivante (crétacé), on n'en connaît qu'une vingtaine. Pourtant, le crétacé a duré 80 millions d'années, les oiseaux ont accompli au cours de ce temps une évolution remarquable, et il a dû en exister un grand nombre et de formes variées. Notre documentation est donc très incomplète.

Ainsi donc les fossiles conservés ne donnent pas une idée précise du groupement vivant. Les oiseaux et les insectes fossilisés dans cette sorte de boue corallienne des calcaires lithographiques de Solenhofen n'ont pas vécu en compagnie des poissons, des crustacés et des méduses. Le vent les a entraînés et ils sont « tombés en mer ». On y trouve aussi des reptiles volants. Mais les reptiles terrestres manquent. Nous avons donc affaire à une population non représentative, à un groupement d'animaux réunis par la mort et qui ne donnent pas une image fidèle de ceux qui ont vécu ensemble et à l'endroit même de leur fossilisation.

C'est encore un classement, une sélection par transport qui se trouve à l'origine de la constitution différentielle des gisements de houille. Car c'est le vent qui a classé les pollens ou spores et les a associés parfois aux algues. Et c'est l'eau qui a flotté les feuillages, les branches et les troncs d'arbres pour constituer les dépôts organiques se trouvant à l'origine des différentes sortes de houille.

Car l'eau est un autre agent de sélection et de classement. Les coquillages vides sont brassés par les vagues, détruits en grande partie et, finalement, rejetés sur les rivages et classés par catégories.

Enfin, le collectionneur lui-même exerce un choix et intervient ainsi pour fausser l'image du monde qu'il veut faire connaître. L'amateur ne cherche que les belles pièces, il dédaigne les petites ou ne s'intéresse qu'à certaines catégories, tandis qu'il faudrait s'efforcer de tout récolter à l'intérieur d'un espace délimité. Rien n'est à dédaigner pour qui veut rétablir l'association réelle, la population d'une parcelle délimitée du milieu biologique. C'est justement le but de la paléoécologie (du grec *palaios*, « ancien », *oikos*, « maison », « demeure », et *logos*, « conception », « doctrine »), science qui cherche à définir les rapports entre le milieu géologique et le monde des organismes qui l'habitent.

C'est là un problème tellement complexe que le chercheur placé en face de si nombreuses difficultés a tendance à le simplifier en exerçant un choix parmi les documents.

Cette « sélection observationnelle » est un fait d'autant plus fréquent que la plupart des chercheurs sont des spécialistes et portent leur intérêt de préférence sur les groupes faisant l'objet de leur étude.

Ainsi, après les chances diverses de préservation, la sélection par transport, les possibilités variables de fossilisation et les hasards de la découverte et de la mise au jour, le savant lui-même exerce encore un choix supplémentaire, une sélection de spécialiste de laquelle résulte souvent une image très imparfaite de la réalité du monde vivant.

Cette courte énumération suffit à montrer l'importance et la difficulté d'obtenir en paléontologie un échantillonnage correct, que l'on rapprochera de celles qui concernent les sondages d'opinion.

la sélection observationnelle en astronomie

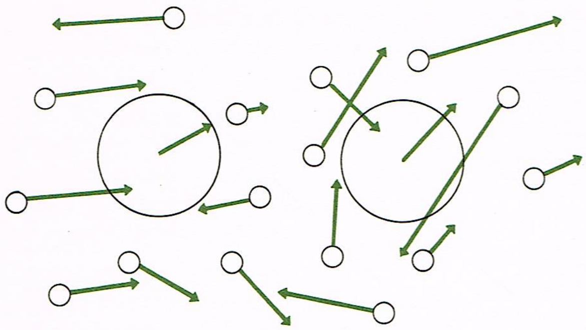
Terminons par un dernier cas révélateur, que nous emprunterons à l'astronomie, où le danger est peut-être plus subtil encore.

Il est très important de connaître les propriétés cinématiques des étoiles, dont les écarts par rapport à un type spectral fournissent des renseignements précieux sur l'évolution des systèmes stellaires (voir le chapitre de l'astronomie dans le premier tome).

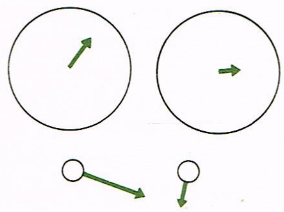
Il faut donc trier d'abord les étoiles par types, et étudier par exemple la vitesse moyenne des étoiles *d*G1 (type solaire).

Mais le nombre de *d*G1 peuplant notre Galaxie est de plusieurs milliards, et l'on n'en connaît bien que quelques spécimens, proches et bien étudiés (une centaine environ). De plus, rien ne prouve que les étoiles proches ont les mêmes propriétés que celles du centre de la Galaxie; et l'astrophysique affirme même très exactement le contraire. L'étude des *d*G1 proches ne peut renseigner que sur les *d*G1 proches. C'est aussi évident que le fait qu'étudier des singes ne peut guère renseigner... sur les ours. *Mais les statistiques ne sont que bon sens; ce qu'on a trop tendance à souvent oublier.*

Ce n'est pas tout, et il y a plus grave : les étoiles du type solaire n'ont pas



La population des dGI proches : une majorité d'étoiles légères, donc de grande vitesse moyenne (équipartition de l'énergie).



La population des catalogues : on est amené à sous-estimer fortement la vitesse moyenne.

toutes le même éclat. Une étoile 4 fois plus brillante qu'une autre s'observe — toutes choses égales — 2 fois plus loin, donc dans un volume $2^3 = 8$ fois plus grand. Nos catalogues comporteront 8 fois plus d'étoiles très brillantes que la stricte équité ne le permettrait. Et l'on observe finalement les propriétés des étoiles les plus brillantes, et non celles de l'astre qui répond aux conditions moyennes. Pour peu que la répartition des vitesses varie avec l'éclat, on sera victime de la « sélection observationnelle », et les conclusions seront erronées; appliquer un terme correctif (pondération) aux éclats d'étoiles, pour tenir compte du choix inconsciemment fait, a conduit à augmenter de près de 20 p. 100 les distances des astres les plus voisins.

une information toujours limitée

Même si les pièges de la sélection ont été décemment évités, et si l'on peut admettre qu'échantillon et population parente ont les mêmes propriétés, il est cependant clair qu'une incertitude demeure, comme on l'a vu lors de l'estimation d'une probabilité par une fréquence.

On démontre très précisément que la dispersion du résultat obtenu (son incertitude) est limitée inférieurement par une quantité ne dépendant que de la forme de la population parente et de la taille de l'échantillon, mais non des résultats obtenus (1) :

$$\sigma^2 \geqslant \frac{1}{nI}.$$

(I est appelé « quantité d'information portée par un tirage sur la population parente », bien qu'il puisse y avoir confusion entre ce terme et celui dont il a longuement été question au chapitre précédent. Les théoriciens discutent encore sur le « joint » exact entre les deux théories qui limitent — pour des raisons différentes — nos connaissances.) On est amené à définir la précision comme l'inverse du carré de la dispersion, et le résultat, finalement, peut s'énoncer

$$p \leqslant nI.$$

En aucun cas la précision ne peut être supérieure au produit nI. Cette dernière borne est proportionnelle au nombre de tirages effectués (supposés indépendants).

C'est ce que nous avons déjà vu à deux reprises, et qui devient maintenant un théorème très général.

(1) La formule suivante (due à Georges Darmois, à Cramer, à Fréchet et à Rao, quatre scientifiques contemporains dont elle porte le nom) suppose les tirages indépendants. S'il n'en est pas ainsi, la quantité d'information est encore plus sévèrement limitée.

l'algèbre des aléas

L'introduction des variables aléatoires fit faire à la science un progrès décisif. C'est, si l'on veut, l'introduction de l'algèbre dans les probabilités, qui, auparavant, n'étaient qu'une science arithmétique de dénombrement. Une variable aléatoire est un ensemble de valeurs auxquelles sont attachées des probabilités.

Une variable susceptible d'un seul état (avec la probabilité unité) est une variable certaine (ou une variable au sens classique du terme).

Le montant x de la face d'un dé est une variable aléatoire $\{X\}$ discontinue pouvant prendre les valeurs entières 1 à 6 avec des chances égales (à 1/6) pour chacune d'elles. On voit que $\{X\}$ est déterminé par une épreuve : le jet de dé.

La taille d'un conscrit de la classe 64 est également une variable aléatoire, pratiquement contenue entre deux limites extrêmes (par exemple 1,45 m et 2,10 m). Mais sa densité de probabilité n'est pas constante, car il y a davantage d'individus de 1,75 m que de 1,50 m.

S'il est intéressant de connaître la répartition des conscrits par tailles, il ne l'est pas moins de savoir combien ont une taille inférieure à 1,60 m, à 1,65 m, etc., bref à x mètres. Cette nouvelle distribution est appelée « fonction de répartition » de la variable $\{X\}$. L'ensemble des individus dont la taille ne dépasse pas 1,70 m est constitué de ceux qui mesurent 1,69 m, 1,68 m, etc. La fonction de répartition est donc construite à partir de la densité de probabilité par addition, ou, comme disent les mathématiciens, par intégration.

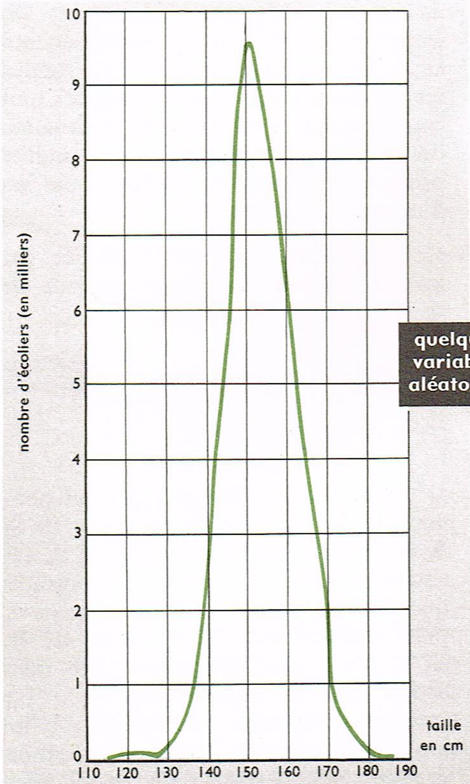


Cette tranche spéciale de la Loterie nationale comportait :

1 lot de	1 000 000 F	1 000 000 F
48 lots de	10 000 F	480 000 F
. . . . lots de	. . . .	. . . . . .
. . . . lots de	. . . .	. . . . . .
20 000 lots de	100 F	2 000 000 F
40 000 lots de	60 F	2 400 000 F
espérance mathématique totale pour 400 000 billets		13 200 000 F

Pour 1 billet $\frac{13\,200\,000}{400\,000} = 33$ F

L' « espérance mathématique » attachée à un billet valant 55 F était de 33 F. En moyenne, on ne pouvait que perdre 22 F et seul l'attrait du risque et du jeu, et une habile publicité peuvent nous conduire à un achat correspondant à une perte (moyenne) de 40 p. 100 de la mise.



Répartition de la taille de 67 000 écoliers de 14 ans en 1950.

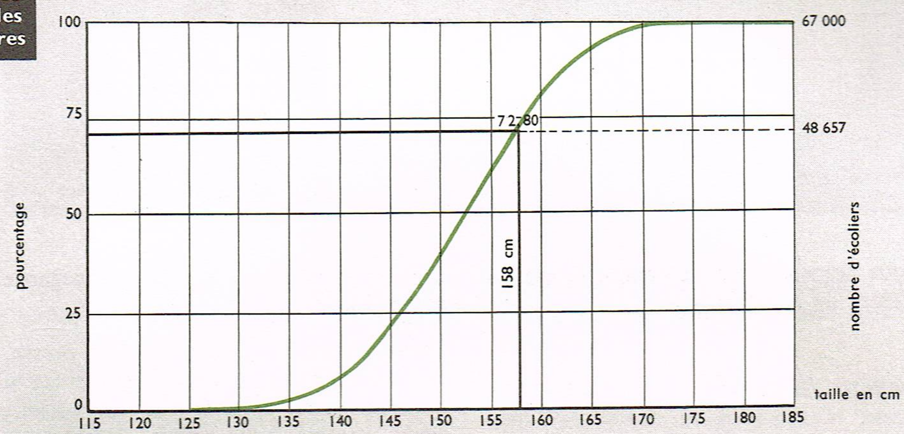
quelques variables aléatoires



une variable aléatoire $\{x\}$, ses valeurs, les probabilités correspondantes, la moyenne et le carré de la dispersion valent :

$$E(x) = \frac{1}{6}1 + \frac{1}{6}2 + \frac{1}{6}3 + \frac{1}{6}4 + \frac{1}{6}5 + \frac{1}{6}6 = \frac{21}{6} = 3{,}5$$

$$\sigma^2 = \frac{1}{6}\left(1 - \frac{7}{2}\right)^2 + \frac{1}{6}\left(2 - \frac{7}{2}\right)^2 + \frac{1}{6}\left(3 - \frac{7}{2}\right)^2 + \frac{1}{6}\left(4 - \frac{7}{2}\right)^2 + \frac{1}{6}\left(5 - \frac{7}{2}\right)^2 + \frac{1}{6}\left(6 - \frac{7}{2}\right)^2 = 2{,}92...$$



Polygone cumulatif de la distribution de la taille de 67 000 écoliers. Les 48 657 écoliers dont la taille est inférieure à 158 cm représentent 72,8 p. 100 du total.

l'espérance mathématique, mesure du comportement moyen

Il existe aussi une grandeur appelée du beau nom d'« espérance mathématique », qui n'est autre que la valeur probable ou encore la moyenne $\bar{x}$. Pour une variable (au sens de l'algèbre) qui prend les valeurs x_1, x_2... x_n, la moyenne est tout simplement la somme

$$x_1 + x_2 ... + x_n$$

divisée par n. Dans le cas d'une variable aléatoire, on pondère les valeurs que peut prendre $\{X\}$ par les chances qu'elles ont de se produire. Ainsi, la valeur probable de $\{X\}$ qui a la probabilité p_1 d'être égale à x_1, p_2 d'être égale à x_2, etc., est $\bar{x} = p_1x_1 + p_2x_2 + ...$ (puisque $p_1 + p_2 + ... = 1$).

On avantage ainsi les valeurs qui ont de fortes probabilités.

Jouons avec 2 dés. Quelle est la valeur moyenne du total que l'on peut espérer? Les valeurs possibles de $\{X\}$ vont de 2 à 12, mais elles n'ont pas la même probabilité, puisqu'il n'existe qu'une possibilité (6 + 6) pour 12 et cinq (6 + 2, ...) pour 8. $\{X\}$ n'est pas équiréparti.

Il est facile de dresser un tableau des probabilités :

Nombre de points	2	3	4	5	6	7	8	9	10	11	12
Nombre de chances sur 36	1	2	3	4	5	6	5	4	3	2	1

Le nombre moyen de points est

$$2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + \cdots$$
$$11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = 7.$$

Beaucoup de lecteurs auraient sans doute parié pour 6 !

la dispersion, mesure de l'individualité

L'espérance mathématique — ou moyenne $(\bar{x})$ — traduit une valeur moyenne de l'ensemble $\{X\}$. Tous les conscrits n'ont pas 1,70 m, et des variations en plus ou en moins ont lieu autour de cette valeur moyenne. Les mathématiciens ont trouvé commode de mesurer cet écart par la dispersion σ :

$$\sigma^2 = \text{valeur moyenne de } (x - \bar{x})^2$$

(il fallait bien prendre le carré pour que l'écart soit toujours positif ou nul). Une dispersion nulle correspond évidemment à une variable certaine $(x = x_0)$.

Il est clair que plus σ est grand, plus probables sont des valeurs très différentes de $\bar{x}$, plus il y a de « jeu », plus chaque individu peut s'écarter du comportement moyen. $x = x_0 \pm \sigma$ signifie que $\{X\}$ est une variable aléatoire de valeur moyenne x_0 et de dispersion σ.

Notons qu'un tel symbolisme et la notion même de dispersion peuvent être ambigus. Quand on dit que la taille des Français est

$$t = 1{,}70 \pm 0{,}05,$$

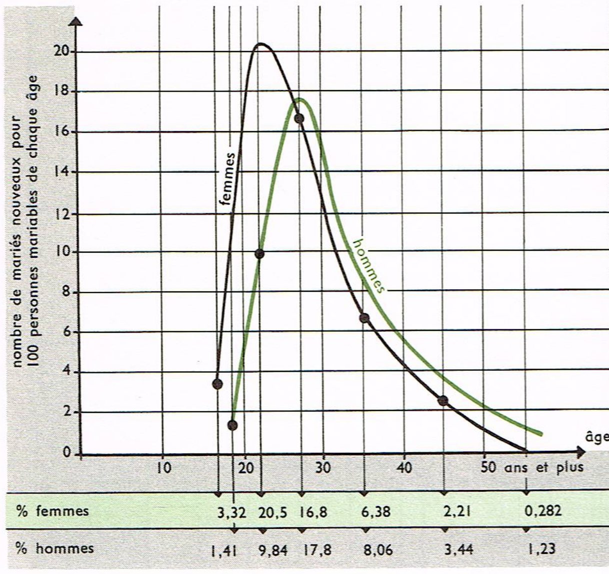
il est clair que tout les Français ne sont pas de la même taille, puisque les mesures sont effectuées avec une précision de 1 ou 2 cm, qui ne peut expliquer des tailles telles que 1,50 m ou 2,07 m. Les Français n'ont pas tous les mêmes caractéristiques, ce qui permet d'ailleurs de les reconnaître.

Au contraire, quand on dit que la charge de l'électron est

$$e = (1{,}602\,06 \pm 0{,}005)\ 10^{-19} \text{ coulomb},$$

la dispersion des résultats peut s'expliquer par l'imprécision des mesures : on n'a jamais pu mettre en évidence de différence significative, c'est-à-dire supérieure aux erreurs de mesure.

Les électrons sont tous identiques (ce qui est normal pour une particule élémentaire) : ils n'ont aucune individualité.





Taux de nuptialité en France
Dois-je passer sous les fourches caudines du mariage?

Nous verrons effectivement que rien ne permet de les distinguer.

suis-je libre de me marier ?

Les statistiques nous révèlent avec une approximation assez grande le nombre des Français célibataires qui vont se marier dans l'année à venir (taux de nuptialité constant, ou du moins très peu variable, et variant de manière régulière et prévisible).

Ne puis-je faire mentir la statistique en refusant de me marier? Mieux encore, en patronnant une campagne de presse bien orchestrée en faveur du célibat, empêchant peut-être des milliers de Français et de Françaises d'unir leurs destinées.

Ne comptons pas sur la première méthode, car ce pourcentage, provenant d'une estimation à partir de données connues (taux de nuptialité des années précédentes, renseignements sur les facteurs susceptibles de le faire évoluer), est pourvu d'une incertitude : s'il est vrai que je peux modifier d'une unité le nombre de mariages en refusant de m'enchaîner par les liens nuptiaux, cela ne peut changer significativement un pourcentage portant sur des millions de cas; tout au plus puis-je espérer le modifier de quelques millionièmes, variation infime et sans grande signification.

De plus, si mon comportement est très incertain, il n'en est pas de même du comportement moyen d'un grand nombre d'individus, pour qui joue la loi des grandes séries. C'est précisément ce qui sauvegarde notre liberté.

Quant à lancer une campagne de presse, et en admettant qu'elle réussisse (et il faudrait alors approfondir les raisons ayant motivé son succès même), et réussisse si triomphalement que s'effondre le taux de nuptialité, c'est une autre affaire. Le futur est toujours incertain, et toute prévision obtenue en extrapolant le passé est sujette à erreur, surtout lorsque intervient un événement nouveau et qu'on n'a pas pu ou su prévoir : celui même de la campagne de presse; tout au plus eussions-nous pu soupçonner qu'une quelconque variation brusque, l'an dernier, du nombre de mariages allait se traduire par un choc en retour, qui favoriserait mon action « préventive » et en assurerait le succès. Il est alors plus facile de prévoir le résultat que de le provoquer à dessein.

la corrélation, mesure de l'interdépendance dans l'indépendance

La corrélation est une notion très utile en probabilités. Elle exprime que deux variables aléatoires ne sont pas indépendantes l'une de l'autre, mais peuvent être liées par des relations plus ou moins strictes. Il est ainsi d'un grand intérêt, pour la prévision, de savoir comment la récolte de blé peut dépendre des températures du mois de janvier, la qualité du vin du soleil du mois d'août, la consommation publique de l'inflation. Les probabilistes utilisent un coefficient de corrélation, qui est nul lorsque les phénomènes sont totalement indépendants. On trouve un coefficient de corrélation nul entre les phases de la Lune et le temps (ce qui détruit une croyance tenace), tout comme entre le caractère et la situation des planètes (horoscope), cette nullité prouvée n'empêchant d'ailleurs pas un florissant commerce.

La corrélation est surtout utilisée par les statisticiens; c'est une méthode de détermination des causes.

la difficile recherche des causes

Il semble facile d'affirmer que tel phénomène A est cause de tel autre B. Si A et B sont seuls en jeu, et si A est antérieur à B (1), il suffit de constater que, dès l'apparition de A, B s'est ensuivi, pour déduire que A est cause de B. Et il n'est guère discutable que mon action sur tel bouton électrique est cause de l'extinction de telle lampe (surtout si le bouton porte la mention « extinction »). Même un singe, libre de faire des milliers de fois des expériences analogues — presser sur un bouton et voir s'éteindre ou s'allumer un signal —, finit par établir un rapport entre les deux faits et pose l'un comme cause de l'autre.

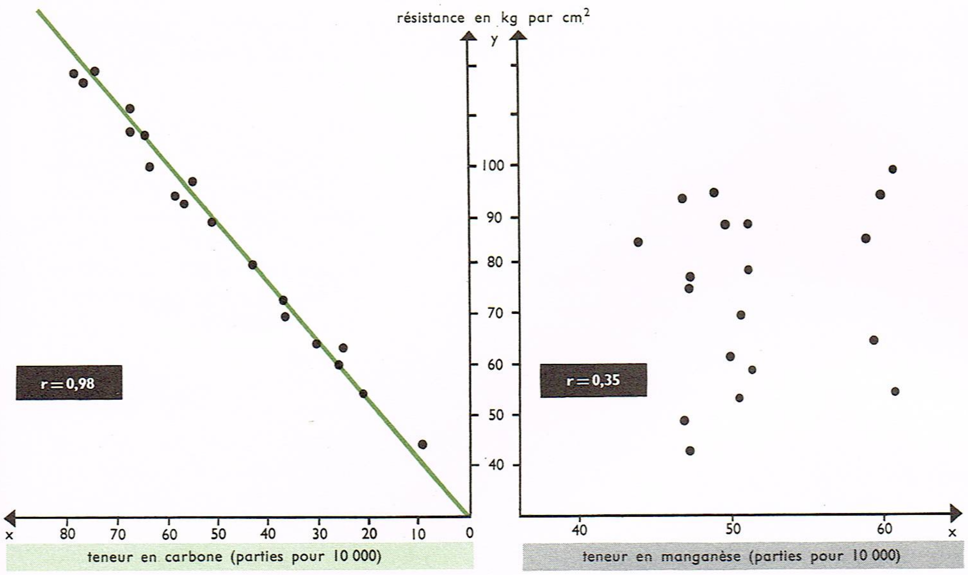
Mais si les phénomènes sont plus complexes, et liés l'un à l'autre? Le temps froid qui a régné pendant l'hiver 1962-63 a-t-il eu pour conséquence une récolte désastreuse des artichauts, l'été suivant? Le renforcement des contrôles routiers a-t-il été cause de la diminution du taux de la mortalité sur les routes? Les expériences atomiques multipliées sont-elles cause du recul des glaciers? Il est facile de répondre à la dernière question : la fonte des glaciers a commencé bien avant 1950. De plus, il semble difficile de trouver quel processus ont pu déclencher les explosions, capable de faire fondre des masses énormes de glace. *Un peu de bon sens — la qualité la plus utile aux statisticiens — suffit pour répondre ici.*

En revanche, il est plus délicat de répondre à la première question : il est certes de simple bon sens que la rigueur de l'hiver a joué, mais d'autres facteurs ont pu intervenir; la sécheresse n'a-t-elle pas eu un rôle encore plus néfaste?

---

(1) La question de l'antériorité pose un problème délicat, dont il a été question en détail dans le chapitre sur la relativité.

Phot. Belzeaux.



Ces deux diagrammes sont relatifs à 19 échantillons d'acier : on porte en ordonnée leur résistance à la traction, en abscisse leur teneur en carbone et en manganèse. Il est évident qu'il y a un rapport étroit entre teneur en carbone et résistance, mais beaucoup plus lâche entre teneur en manganèse et résistance.
On précise, en calculant le coefficient de corrélation r :

$$r = \frac{n\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)}{\sqrt{n\Sigma x_i^2 - (\Sigma x_i)^2}\ \sqrt{n\Sigma y_i^2 - (\Sigma y_i)^2}},$$

qui est toujours compris entre — 1 et + 1, $|r| = 1$ entraînant que les points sont tous alignés sur une droite;
$|r|$ voisin de 1, que les points sont presque alignés : l'ordonnée dépend fortement (et linéairement) de l'abscisse ;
$|r|$ petit (figure de droite) : ordonnée et abscisse sont d'autant plus indépendantes que r est petit.
r mesure en quelque sorte la dépendance (linéaire).

On ne peut y répondre (sauf en faisant de nombreuses expériences au phytotron de Gif-sur-Yvette, créé précisément à coups de millions pour permettre une réponse expérimentale directe) en faisant varier le seul climat de janvier : il faudrait se placer dans un monde où l'on serait maître de renouveler indéfiniment, ou de modifier à sa guise, un processus aussi lent et complexe que la croissance et la récolte de millions d'artichauts.

Un coup d'œil sur les figures suffirait à montrer dans quelle direction est la réponse. **Un grand coefficient de corrélation signifie : rapport.** Plus exactement : on est amené à soupçonner que deux faits sont en rapport (en corrélation) lorsque le coefficient (précisément appelé « de corrélation ») est significativement non nul.

est-il dangereux d'aller voir le médecin ?

Attention! Rapport ne signifie pas nécessairement relation de cause à effet. De ce fait, trop évident, que la probabilité de mourir est grande chaque fois que l'on a vu un médecin dans la semaine — plus généralement, qu'il y a une indéniable et mesurable corrélation entre visite médicale et mort prochaine —, gardons-nous de déduire, contrairement à tout bon sens, que les médecins tuent.

Mais soupçonnons plutôt une cause commune. Et pensons tout de suite à la maladie, qui a entraîné à la fois la nécessité d'une consultation et une issue fatale (si ce raisonnement ne nous suffit pas, calculons la corrélation qu'il y a entre maladie et visite médicale, et celle qui existe entre maladie et mort; les deux sont beaucoup plus marquées que celle qui nous a donné l'éveil).

l'astrologie a-t-elle une base scientifique ?

La croyance en une influence de la position des astres sur les destinées humaines, qui est enrichissante pour ceux qui font profession d'y croire, a une origine immémoriale. Il n'est pourtant guère besoin d'être grand clerc pour affirmer qu'il n'est guère probable qu'un jour on trouve un processus expliquant comment la position de α de la Grande Ourse, étoile dont la distance s'exprime en milliers de milliards de kilomètres, a pu contraindre les gènes et chromosomes des Terriens à ne produire que des descendants guerriers, ou, au contraire, rien que des pacifiques.

De nombreuses études de corrélation ont été faites, et toujours le résultat fut négatif.

Une seule explication subsiste pour un engouement aussi peu plausible : est-il possible de naître plus exactement sous le même signe que des jumeaux? et peut-on espérer corrélation plus étroite que celle de leurs caractères? Même si ce phénomène s'explique beaucoup plus naturellement par l'existence de la gémellité véritable (que ne pouvaient suspecter les Chaldéens) que par l'identité des influences astrales, il n'en a pas fallu plus pour que naquît l'astrologie (notons que l'on eût tout aussi bien pu admettre que les caractères dépendent du lieu de naissance, deux vrais jumeaux naissant non seulement à la même heure,



Phot. Alinari.

mais aussi dans la même chambre : on aurait fait non de l'astrologie, mais de la « géologie », en un sens de ce terme bien différent du sens courant).

Et, sur cette minime coïncidence, les hommes, ivres de rêve, avides d'attacher de l'importance aux événements favorables à leur thèse (en rejetant systématiquement dans l'oubli les événements défavorables, fussent-ils immensément plus nombreux), anxieux de se livrer aux sortilèges de la magie pour connaître l'avenir (même si la simple raison est une arme plus certaine pour y parvenir), ont perpétué une croyance, si contraire à l'expérience du simple bon sens, mais si conforme aux désirs les plus profonds de leur subconscient.

l'éternel retour des cataclysmes

« Il n'y a plus de saison; pas étonnant, avec leurs inventions de la radio » (1910). « Encore un hiver rigoureux, c'est « leur bombe » (1960).

On pourrait multiplier les exemples. Tout, même le climat, va-t-il de mal en pis? Ou bien l'homme est-il sensible aux mauvaises périodes, mais beaucoup plus distrait quand le temps est beau (à son sens du moins; l'opinion des citadins et des campagnards peut différer sur ce point, ou celle d'un Africain et d'un Suédois, tous deux à Paris au mois de mai)?

Là encore le scientifique, sous l'accumulation de slogans du type précédent, a cru devoir vérifier l'évolution du climat en fonction du temps, faisant ce que l'on appelle, un peu pompeusement, de l'*analyse chronologique*, ou encore de l'« analyse des séries chronologiques ». Les calculs ont confirmé ce qui apparaît déjà clairement dans les courbes publiées ici : les cataclysmes ne sont pas l'apanage de notre génération. Le climat ne se détériore pas; ce sont nos souvenirs qui enjolivent le temps de notre jeunesse. En revanche, l'analyse a révélé très nettement l'existence d'une période de onze ans et demi dans les conditions météorologiques. Comme, d'autre part, les astronomes avaient précisément découvert une période égale dans l'agitation solaire, ils n'eurent pas grand mal à associer deux faits aussi liés et dont l'un pouvait directement expliquer l'autre quand on sait que finalement toute énergie nous vient du Soleil.

les schémas de l'évolution

Un point est de connaître un phénomène aléatoire à un instant donné; un autre, de prévoir son évolution. C'est le difficile problème de l'évolution d'un phénomène aléatoire au cours du temps, ce que nous venons d'appeler l' « analyse des séries chronologiques ». La règle à laquelle le phénomène obéit lorsque le temps s'écoule se nomme « processus ». Il y a, bien sûr, de nombreuses classes de processus.

Un cas est celui où la chaîne des événements est faite de maillons indépendants, telle la suite des couleurs tirées à la roulette, où chaque coup est indépendant des précédents, telle la réception des appels sur un central téléphonique, telle l'émission des électrons d'une cathode de lampe de radio, où l'éjection d'un électron ne dépend nullement de celle qui a eu lieu un instant auparavant. Ce sont des *processus de Poisson*. Outre leur grande fréquence dans les phénomènes physiques, ces processus présentent ce grand intérêt qu'ils caractérisent les phénomènes qui ne vieillissent pas, car ils définissent ceux dont les chances d'apparition à un instant donné ne dépendent pas des instants précédents (transmutation d'un radio-élément). Ces phénomènes n'ont pas d'histoire (mathématiquement, cela s'exprime par le fait que les équations qui représentent les lois sont du type exponentiel). On rencontrera donc très peu de processus de Poisson en biologie.

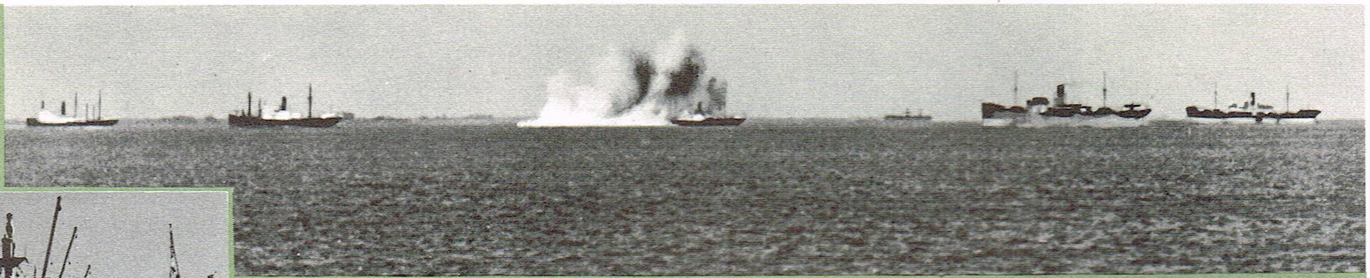
Tous les processus non formés d'enchaînements indépendants appartiennent à d'autres types. Tels sont les processus de Markov, qui s'appliquent au cas où l'état probable d'un système dépend de son état immédiatement antérieur, mais non du passé en général. Les *processus de Markov* règlent l'évolution des événements à mémoire courte. Par exemple, un joueur joue à la roulette sur les couleurs; si c'est un joueur sans préventions, à chaque noir sorti il jouera un jeton sur le rouge ou le noir sans tenir compte du résultat du tirage précédent. La suite de ses mises obéit à un processus de Poisson. Mais si l'homme est un peu superstitieux, et qu'il joue chaque fois la couleur qui n'est pas sortie la fois précédente, le processus devient markovien.

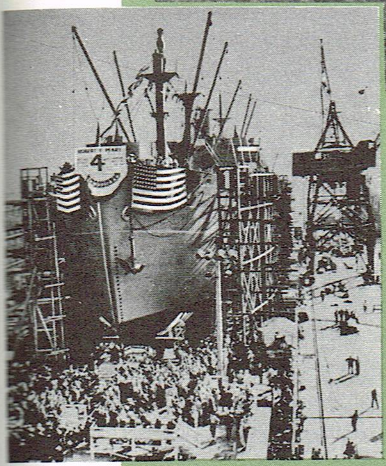
Les phénomènes de diffusion sont en général de ce type. Tel est le cas de la goutte de vin précautionneusement versée dans un verre d'eau. Il est clair qu'au moins au début, quand la concentration en vin reste localement élevée, la distribution de « particules » de vin dans le mélange est très tributaire de leur état précédent. Au fur et à mesure que le temps s'écoule, le mélange s'homogénéise; le lien causal entre présent et futur devient plus lâche, jusqu'au mélange parfait, sans histoire, obéissant au processus de Poisson.

Paris, 1910. Ce n'était pourtant pas la faute de la bombe atomique... ni même celle de la télégraphie sans fil.



Phot. Neurdein.





LA RECHERCHE OPÉRATIONNELLE

Il s'agissait de savoir — lors de la Seconde Guerre mondiale — s'il était possible de surmonter les dramatiques torpillages de convois dans l'Atlantique, et de connaître, par exemple, le tonnage de navires à lancer chaque année pour combler les pertes. S'il dépassait les possibilités des chantiers de construction, la bataille eût été perdue. Fallait-il mieux, en outre, disperser les bâtiments ou les disposer en convois sous escorte? Intervenaient le nombre S de sous-marins allemands, le nombre *n* de cargos rescapés à acheminer coûte que coûte entre les États-Unis et l'Angleterre, le nombre N de cargos devant appareiller pour donner *n* rescapés, le nombre t de torpilles par sous-marin.

Puis les probabilités suivantes (d'estimation délicate et quelquefois très hasardeuse) :

P_1 pour un U-Boot d'apercevoir un cargo isolé;

P_2 d'apercevoir un convoi;

P_3 (*t*) de couler le cargo isolé, compte tenu du nombre t de torpilles;

P_4 (*t*, *c*) de couler *c* cargos du convoi;

P_5 pour le sous-marin d'être coulé après avoir repéré un cargo isolé;

P_6 pour ce même submersible d'être coulé par l'escorte d'un convoi.

En appliquant les lois des probabilités, on en déduisait le nombre de navires arrivant à bon port, le nombre de navires coulés et, par suite, celui des cargos à mettre en chantier chaque année.

On en a conclu :

1° Qu'il était préférable de grouper les navires en convois;

2° Que la bataille pouvait être gagnée — ce qui fut fait. Il faut dire que les hypothèses étaient solidement fondées sur les comptes d'exploitation de la marine et que le nombre de navires antérieurement envoyés par le fond était suffisant pour que l'outil probabiliste fût employé avec succès.

Depuis ce premier succès, la recherche opérationnelle a étendu son domaine d'application à toutes les disciplines; citons par exemple les « jeux » qui consistent à simuler la gestion d'une entreprise suivant plusieurs prévisions et éventualités, ainsi que la théorie des queues (ou l'art, utile, de savoir organiser une file d'attente).

le prix d'une vie humaine

Une question délicate, tant elle comporte d'implications personnelles et psychologiques, se pose aussi, et nous terminerons par elle. Si l'on doit pondérer tous les risques, et les évaluer même symboliquement en termes d'argent — ou mieux de temps et d'efforts —, quelle valeur attribuer à la vie humaine? Peut-on la chiffrer, et comment?

Il m'est agréable et utile de traverser la rue pour acheter le journal auquel je suis habitué; la lecture des bandes dessinées me distraira un peu, la page politique va peut-être me permettre un vote plus raisonné, et celle de la page boursière un arbitrage fructueux. Mais, si l'on admet que l'insécurité du piéton est plus grande sur la chaussée — même sur les passages cloutés — que sur le trottoir, qui lui est encore largement réservé, je prends un risque.

Ce risque, nous n'avons guère l'habitude de le chiffrer — et il est probable que cela n'en vaut que mieux; mais il est facile de trouver des cas, plus réels et plus complexes, où la question peut se poser.

Il n'est guère sensé de répondre que le prix de la vie est infini, car c'est le bien le plus précieux que l'on possède, et il conditionne les autres. Une telle réponse nous interdirait de traverser la rue pour lire le journal. Elle nous interdirait toute action, aboutissant ainsi à la plus stérile des inerties. La vie « dans le coton » n'est guère tentante et ne vaut guère d'être vécue. Elle n'est même guère possible; peut-être même est-ce finalement la plus dangereuse, une prophylaxie excessive nous laissant démunis en face de la moindre attaque.

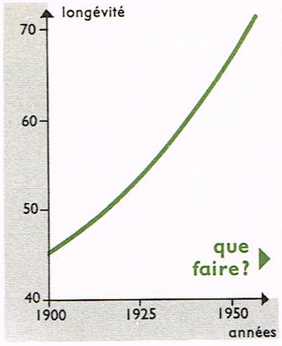
Comment donc l'estimer et à combien? Les juges le tentent, en fixant les dommages-intérêts dus à la suite d'un homicide par imprudence, tenant même compte du prix de la douleur *(pretium doloris)*. **Mais qui ne voit le caractère légaliste et insuffisant d'une telle évaluation, et qui ne serait prêt, s'il le pouvait, à payer une « rançon » supérieure pour sauvegarder sa vie?** De même l'évaluation des compagnies d'assurance, ne dépendant finalement que des primes payées, est encore plus arbitraire.

Nous butons ici sur le caractère subjectif que chacun attache, à juste raison, à son existence (« guenille si l'on veut, ma guenille m'est chère »).

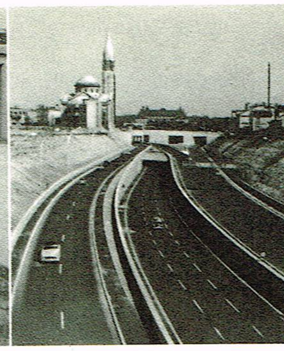
Il est peut-être plus simple de fixer une valeur à la vie des êtres qui ne nous sont psychologiquement rien — ou bien peu. Le gouvernement est-il en droit de raisonner ainsi : « Une autoroute coûte tant le kilomètre; elle économise tant de vies humaines, c'est-à-dire qu'elle conduit à une telle économie (1 vie = 210 000 francs), à ajouter à celle de pneus, d'essence, de matériel, de temps, d'événements... Donc elle est rentable. »

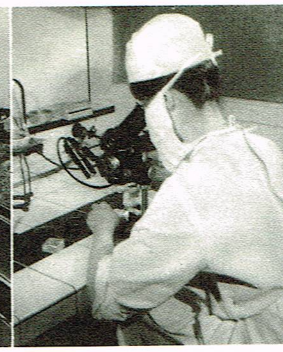
C'est déjà plus raisonnable, et les économistes ne se privent pas d'employer de tels arguments. Même ainsi, il y a des difficultés : la valeur « technique » d'une vie est fonction, entre autres, du revenu annuel. La vie du Parisien moyen est alors plus précieuse (au sens économique du terme : à évaluer plus chère) que

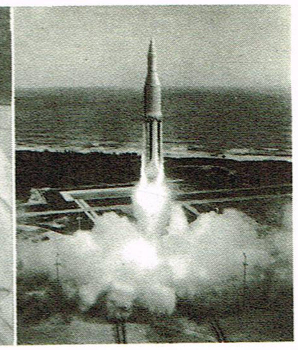
Phot. U.S.I.S. et A.F.P.











celle du Corrézien, en moyenne plus pauvre et moins « rentable ». **Mais qui osera épouser des vues aussi inhumaines ? Si la vie se ramène à de tels calculs sordides, vaut-il la peine de vivre ?**

Aussi est-il mieux de poser le problème un peu différemment. Les ressources de tout gouvernement sont limitées, et encore plus celles qui sont consacrées à la sauvegarde des vies (beaucoup plus généreusement alloués sont les crédits du ministère de la Guerre). Il faut donc faire des choix. Sauvera-t-on plus de vies humaines en affectant les crédits au renforcement des contrôles policiers sur les routes, ou en multipliant les autoroutes ? en organisant un service de prophylaxie médicale et sociale ou en subventionnant largement les biologistes qui feront reculer les maladies ? ou, enfin, en envoyant une fusée sur la Lune avec l'espoir d'y découvrir des remèdes qu'on n'imaginerait pas sur la Terre ? Le problème peut se traiter plus calmement, en termes de coûts et de risques chiffrables, et l'on peut espérer une réponse raisonnable à une question ainsi posée.

l'art de distribuer des récompenses : les statistiques de Maxwell-Boltzmann, de Bose-Einstein et de Fermi-Dirac

Un chef d'entreprise veut récompenser un, ou plusieurs, de ses trois collaborateurs directs, Pierre (P), Jean (J) et Louis (L). Il dispose de deux témoignages possibles de satisfaction. Combien de possibilités a-t-il de rétribuer ses hommes ?

Le problème n'est pas si simple qu'il le semble, il existe plusieurs solutions, en fonction de la nature des primes à distribuer.

Première statistique : les récompenses sont différentes (discernables) et cumulables. Il s'agit par exemple de deux primes de montant différent
(A = 10 000 F et B = 5 000 F),
et le directeur ne s'interdit pas de les donner toutes deux au même collaborateur (et rien aux deux autres).

Il y a trois possibilités pour A, trois pour B, soit $3 \times 3 = 9$ choix en tout.

1° A à P, B à P aussi;
2° A à P, B à J;
3° A à P, B à L;
4° A à J, B à P (ce qui n'est pas équivalent au second cas, A et B différant);
5° A à J, B à J aussi;
6° A à J, B à L;
7° A à L, B à P;
8° A à L, B à J;
9° A à L, B à L aussi.

Deuxième statistique : les primes sont égales, ou indiscernables (toutes deux égales à 7 500 F par exemple), le cumul étant toujours considéré comme possible. Il est alors indifférent de donner A à P et B à J ou A à J et B à P.

Il reste 6 cas (1°, 2° ou 4° — qui sont identiques —, 3° ou 7°, 5°, 6° ou 8°, 9°), dont 3 où les récompenses se cumulent sur une même tête (respectivement P, J, L), et 3 où les récompenses sont partagées, le troisième (respectivement P, J, L) ne recevant rien.

Troisième statistique : les primes sont indiscernables, mais ne peuvent se cumuler (par exemple, il s'agit de procurer le même échelon d'avancement à deux collaborateurs). Restent alors 3 cas, dans lesquels P, J, L respectivement sont brimés (J, L récompensés ; L, P récompensés ; P, J récompensés).

Quatrième statistique : les récompenses sont discernables, mais non cumulables (on classe un collaborateur premier, un autre deuxième, ce qui est une récompense moindre ; remarquons que le collaborateur restant est nécessairement troisième). Il y a 6 cas (ordre de classement PJ, PL, JP, JL, LP, LJ), puisqu'il faut retrancher de la première statistique les 3 cas de cumul.

Le lecteur pourra s'exercer à rétribuer trois collaborateurs en distribuant trois primes (il trouvera respectivement 27, 10, 1 et 6 possibilités). S'il y a n collaborateurs et autant de primes, la première statistique conduit à n^n choix, la troisième à un seul (chaque collaborateur a une prime, et toutes les primes sont équivalentes), la dernière à $n!$

L'exercice précédent n'est pas un pur jeu d'esprit ; il suffit, pour s'en assurer, de noter les noms que portent les trois premières statistiques (la quatrième n'étant jamais considérée en physique) pour voir que les plus grands noms de la science moderne les ont considérées.

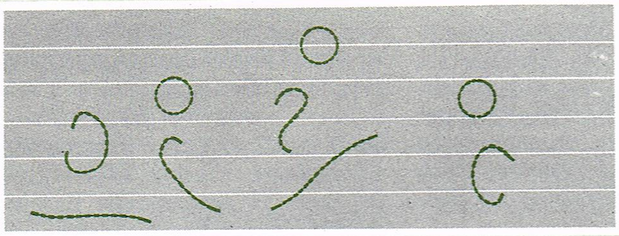
Il est, en effet, essentiel de savoir dénombrer les particules en fonction des états qu'elles occupent, comme nous l'a déjà appris le chapitre précédent et comme nous le montrera encore plus le suivant. Et le résultat est simple à retenir, même s'il ne nous a pas été facile de le dégager : à la différence des assemblages de molécules, d'atomes ou d'ions, aucune particule élémentaire ne suit la première statistique (dite « de Maxwell-Boltzmann », ou encore « classique »), ni la quatrième. **Elles ne suivent que les statistiques les considérant comme indiscernables,** c'est-à-dire celles de Bose-Einstein (possibilité de cumul) ou de Fermi-Dirac (cumul exclu en vertu du principe de Pauli). Nous retrouvons la distinction entre bosons et fermions comme une clé essentielle pour la compréhension du monde moderne, de même que le refus d'obtempérer à des dénombrements « classiques » (indiscernabilité des particules) a été un des premiers glas sonnant la mort de la physique du XIXe siècle (paradoxe de Jeans).

Phot. Kay Lawson, Rapho ; Pougnet, Rapho ; doc. E. Wolff, phot. Larousse ; phot. U.S.I.S.

Quelques jeux

calculer π en lançant des aiguilles

Pouvez-vous calculer π, enfermé dans un bureau au parquet bien régulier (lames de 10 cm), en ne possédant qu'un jeu d'aiguilles de 5 cm (plus généralement de longueur e/2, la moitié de la largeur e des lames)?
C'est très simple, du moins si vous en avez le loisir et le désir : vous lancez votre paquet d'aiguilles au plafond, et n'importe comment. Et vous comptez, après retombée, la proportion des aiguilles qui coupent une rainure : vous estimez ainsi $1/\pi$!
Il existe une démonstration fort ingénieuse du résultat. Supposons qu'au lieu de jeter des épingles vous jetiez une chaîne souple L, formée de maillons. Le nombre de points d'intersection avec les rainures est certainement proportionnel à la longueur de la chaîne ($n = KL$) et ne dépend pas de sa forme. Dans le cas où la chaîne prend la forme d'une circonférence de rayon e/2 ($L = \pi e$), on a sûrement 2 points d'intersection. Par conséquent, $2 = K\pi e$, donc $K = 2/\pi e$.



Appliquons le résultat précédent à notre aiguille ($L = e/2$),

$$n = \frac{2}{\pi e} \cdot \frac{e}{2} = \frac{1}{\pi}.$$

Ainsi, en moyenne, il y a 1 aiguille sur π qui coupe les rainures. L'expérience, bien sûr, fut faite maintes et maintes fois. Elle peut être extrêmement précise (quelques millièmes) si l'on dispose d'un nombre suffisant d'aiguilles. Elle n'a qu'un inconvénient : il faut ensuite ramasser les aiguilles.

le 421

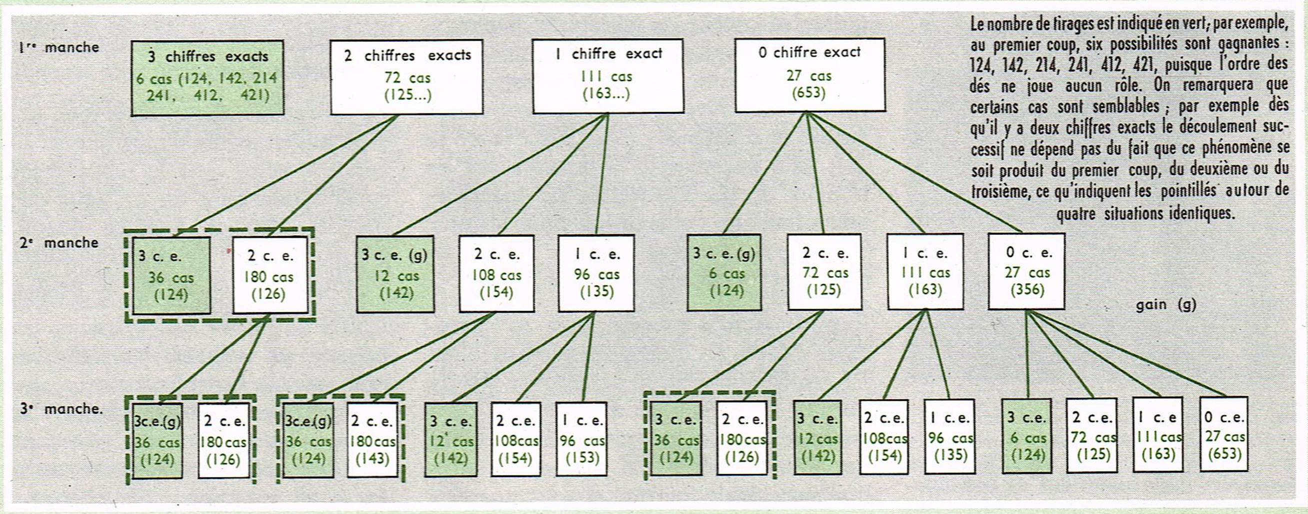
Un peu simplifiée, la règle du jeu est la suivante : on lance trois dés ; le tirage effectué, on peut jeter à nouveau ceux des dés que l'on veut, en nombre quelconque (0, 1, 2, ou même 3). L'obtention du résultat 4-2-1 (quel que soit l'ordre des dés) est seule gagnante.

Quelle est la probabilité de gagner en trois coups au plus? Il va de soi que si le joueur ne réussit pas à abattre 421 du premier coup, il gardera, au cours de la partie, les bons dés, c'est-à-dire ceux marqués 4, 2 ou 1, et ne rejouera que ceux qui lui ont amené 3, 5 ou 6, ou, évidemment, tous les 1, 2 ou 4 en double ou en triple.
Il est commode de s'aider d'un tableau qui fait état des possibilités.

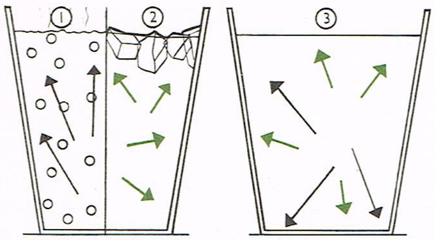
première manche

Il y a $6^3 = 216$ cas possibles, puisque chaque dé peut amener 6 figures (dans ce paragraphe, on tient compte de l'ordre des tirages).
Les cas favorables à l'obtention de 4, 2, 1 du premier coup sont au nombre de $3! = 6$ (124, 142, 214, 241, 412, 421).
Comptons les cas où l'on obtient un seul chiffre bon et deux mauvais, par exemple au moins un 1, mais ni 2 ni 4. Le dénombrement en est facile, pour peu que l'on opère avec soin.

a) Première possibilité : les 3 tirages sont tous 1
(1, 1, 1) *1 cas;*
b) Deux tirages 1, le troisième différent; le troisième ne peut donc être qu'un 3, un 5 ou un 6 (ni 1, ni 2, ni 4) :
(1, 1, 3) (1, 1, 5) (1, 1, 6)
(1, 3, 1) (1, 5, 1) (1, 6, 1)
(3, 1, 1) (5, 1, 1) (6, 1, 1) *9 cas;*
c) Un tirage 1, deux autres différents (ni 1, ni 2, ni 4); 9 tirages où le 1 vient en tête (1, 3, 3), (1, 3, 5), (1, 3, 6), (1, 5, 3), ..., (1, 6, 6) et autant où 1 vient à la deuxième place (3, 1, 3), ..., et autant à la troisième *27 cas.*
Soit $1 + 9 + 27 = 37$ tirages où il faut garder seulement un 1. Autant où il faut garder un 4, autant un 2 (toutes les figures jouent le même rôle), soit 111 cas.
On obtient encore plus facilement le nombre de cas où aucun dé n'est à garder. Il faut que chaque tirage ait donné un des trois mauvais résultats 3, 5, 6. Soit $3^3 = 27$ cas où il faut ne garder aucun dé.
On obtiendrait de même 72 comme nombre de tirages donnant lieu à deux bons résultats et un mauvais. On peut aussi remarquer que $6 + 111 + 27 + 72 = 216$, ce qui permet de vérifier les calculs ou de déterminer une des quatre probabilités connaissant les trois autres.
L'analyse est un peu subtile (il faut dénombrer tous les cas et éviter de compter deux fois le même résultat), mais possible. Le reste du tableau se calcule de même (deuxième manche et troisième). Finalement, la probabilité cherchée est de 0,22 (22 p. 100).

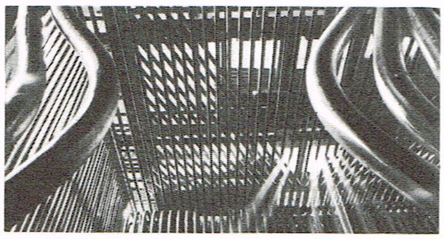


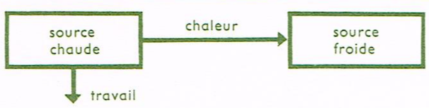
Le nombre de tirages est indiqué en vert; par exemple, au premier coup, six possibilités sont gagnantes : 124, 142, 214, 241, 412, 421, puisque l'ordre des dés ne joue aucun rôle. On remarquera que certains cas sont semblables ; par exemple dès qu'il y a deux chiffres exacts le découlement successif ne dépend pas du fait que ce phénomène se soit produit du premier coup, du deuxième ou du troisième, ce qu'indiquent les pointillés autour de quatre situations identiques.



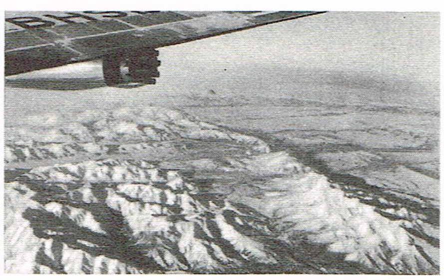
1 : eau chaude (molécules rapides)
2 : glace (molécules à faible vitesse)
3 : un verre d'eau tiède : un état infiniment plus probable (équirépartition des vitesses) en vertu de la loi des grandes séries.

L'Univers évolue vers les systèmes les plus probables.





Le foyer chauffe l'air ambiant, ou celui du condenseur. Les températures s'égalisent, mais du travail est fourni.



Les températures sont basses et égales. Cette vue polaire donne l'idée de ce que pourrait être l'étape précédant la mort thermique de l'Univers.



En fait, la température « zéro absolu » est beaucoup plus basse encore (— 273 °C). Aucun mouvement n'est plus possible. L'Univers est définitivement et inéluctablement figé. Le vase Dewar, dont on voit ici la double paroi, sert à contenir de l'hélium liquide (— 269 °C ou 4 °K).

une science nécessairement probabiliste

Ainsi donc, comme l'a montré le simple exemple du jeu de dés, et comme l'a confirmé le paragraphe consacré à la valeur de la vie, l'homme — composé d'un nombre immense de molécules — intervient de manière si complexe, son action implique tant de phénomènes, qu'il est normal de traiter ses actes en termes de probabilités. Les systèmes physiques peuvent-ils être soumis aux mêmes méthodes ?

Quoi de plus simple que le mouvement d'une particule élémentaire, ou celui d'une étoile ? Peu de forces sont en jeu. L'analyse exacte semble possible, les schémas probabilistes ne s'introduisant que par suite de la difficulté de maîtriser par le calcul des situations trop complexes ou trop mal connues. Nous allons pourtant voir qu'il n'en est rien.

pression et température : des notions statistiques

L'esprit est évidemment impuissant à déterminer, même par le calcul, le mouvement d'une poussière, d'une molécule, d'un électron, ou même d'une étoile parmi des milliards. Non par ignorance, car il connaît les lois des chocs et des attractions, mais la recherche de la solution ne vaut vraiment pas le temps immense qu'elle réclame, ni surtout qu'on lui attache un tel intérêt. Personne ne porte la moindre attention à telle molécule d'azote — en admettant qu'on sache l'identifier — qui se promène à 10 000 m d'altitude; mais Air France se doit de s'assurer que la pression sur les hublots de ses Caravelles est pratiquement toujours constante à cette altitude et n'a pratiquement aucune chance de diminuer brutalement de 50 p. 100.

Rappelons qu'il y a quelque $3 \cdot 10^{21}$ particules d'azote ou d'oxygène par litre d'air. Le probabiliste n'est pas effrayé par ce nombre inimaginable. Au contraire, cette énormité même le tranquillise : la loi des grandes séries lui assure, en effet, que ses calculs, portant sur les nombres astronomiques, ne risquent pas d'être entachés significativement des incertitudes qui pèsent sur le traitement d'un trop petit nombre d'événements. Il applique les lois des chocs aux particules, fait quelques hypothèses simples... et trouve que les vitesses de toutes les molécules de gaz se répartissent autour d'une valeur moyenne u selon une « courbe en cloche ». Bien plus, il trouve que, pour une molécule-gramme M, pression, vitesse et volume moyen sont liés par la relation

$$PV = \frac{2}{3}\left(M\,\frac{u^2}{2}\right).$$

Les physiciens dressent alors l'oreille, car la célèbre loi des gaz parfaits est pratiquement identique :

$$PV = RT_k,$$

où R est la constante universelle, T_k la température absolue ($T_k = T_c + 273$, T_c étant la température Celsius). On trouve donc que

$$\mathbf{Mu^2/2 = 3/2\ RT.}$$

Mais $Mu^2/2$, c'est précisément l'énergie cinétique moyenne du gaz. Ce que nous nommons « température » n'est pas autre chose que l'énergie cinétique moyenne de l'agitation des molécules.

Ainsi nous commençons à connaître les gaz. C'est un fourmillement de molécules tantôt lentes, tantôt très rapides, dont les vitesses obéissent à la loi de Gauss, autour d'une valeur moyenne dont la manifestation sensible est la température. La vitesse moyenne des molécules d'oxygène est par exemple de 420 m/s à la température ordinaire : 15 p. 100 d'entre elles dépassent 630 m/s; 15 p. 100 sont au-dessous de 210.

Le zéro absolu prend alors tout son sens : u est nul en même temps que T. La température zéro, c'est le repos absolu. On conçoit pourquoi on ne peut que l'approcher, car il faudrait refroidir la matière à l'abri de tout rayonnement et de tout choc, ce qu'on ne peut réaliser qu'imparfaitement.

Les corps solides, les cristaux, qui ont une « température », ne sont donc jamais immobiles et les atomes disposés aux sommets des mailles cristallines oscillent autour de leur position d'équilibre. Si la température s'élève trop, l'édifice se rompt et le cristal fond.

Doc. Stein et Roubain, Air France, phot. C. Lacheroy (doc. Air liquide).

Il est un domaine où cette notion, appelée « théorie cinétique des gaz », a besoin d'être fortement amendée, sinon abandonnée. C'est celui des atmosphères raréfiées que l'on trouve dans l'espace, dans les vides poussés que sait faire régner le physicien dans ses instruments, dans les plasmas enfin; la densité n'y est pas suffisante pour qu'on puisse appliquer avec intérêt la statistique et la loi de Gauss. Cette fois, l'importance des effets moyens le cède à celle des effets locaux. L'astronaute, de même, se moque de la pression des météorites, qui, en moyenne, est quasi nulle, mais s'inquiète beaucoup du risque d'en rencontrer une.

le démon de Maxwell

Si les atomes constituant le gaz ont des vitesses différentes, ne peut-on les trier, en séparant les particules rapides de celles qui sont plus lentes ? A la fin de l'opération, une partie de l'enceinte serait occupée par du gaz chaud, et l'autre partie par du gaz froid.

Ce paradoxe, un des plus intéressants de ceux qui furent soulevés par la thermodynamique classique, a été explicité par Maxwell sous la forme suivante. Deux pièces, A et B, remplies d'air à la même température, sont séparées par une paroi. Une minuscule porte, ne laissant passer les molécules qu'une par une, existe cependant; un démon tout aussi minuscule en est le gardien (une opération aussi minutieuse ne déplaît pas nécessairement aux esprits infernaux). Ses consignes sont simples, sa tâche presque routinière : chaque fois qu'une molécule rapide venant de la pièce A demande le passage, il lui ouvre; et de même pour toute particule lente sollicitant le transit de B vers A. Sinon, la porte reste close.

Les démons sont patients. Au bout d'un certain temps, grâce à cette consigne scrupuleusement suivie — un démon n'est-il pas infaillible —, la pièce A se sera appauvrie en particules rapides et enrichie aux dépens de B en molécules lentes. A se sera refroidie, et B réchauffée. **Cela constitue un scandale insupportable en thermodynamique; le second principe est violé** (en absence de tout apport d'énergie, les températures ne peuvent que s'égaliser); Carnot et Clausius se retournent dans leur tombe. Entre les deux pièces à températures désormais différentes, on peut brancher une petite machine à vapeur et produire du travail à partir d'une source de température initialement uniforme. L'anomalie diabolique est de même ordre que celle qu'on ne manquerait pas de constater si, en faisant chauffer une casserole d'eau, on voyait l'eau geler et la flamme devenir encore plus chaude. (Il suffirait pourtant de tendre entre le feu et la casserole un tamis de portes gardées par une légion de démons de Maxwell ne laissant passer en direction de l'eau que les molécules les plus froides de la flamme.) Il va de soi que le paradoxe du « démon de Maxwell » est critiquable sur bien des points. Il est notamment difficile d'admettre que la manœuvre d'une porte ne consomme aucune énergie. On a aussi fait observer que le diable ne pouvait pas voir les molécules. Quoi qu'il en soit, et si l'on décide que l'ingéniosité des démons touche à la perfection, il reste une condition absolument nécessaire de l'expérience : pour ouvrir ou fermer sa porte, le diable doit savoir si la molécule qui frappe est plus rapide ou moins rapide que la moyenne. Les molécules ne portent pas cette qualification sur elles; il faut donc mesurer les vitesses. Nous, gens du XXe siècle, nous dirons que le démon doit être informé. **La théorie de l'information semble ainsi venir au secours de la thermodynamique en difficulté. Qu'en est-il au juste ?**

des lois de la thermodynamique à celles de la vie

Si un message peut revêtir N formes également probables, on a vu que l'information qu'il contient est $I = \log_2 N$, que l'on peut encore écrire

$$\mathbf{I = -\log_2 p}$$

en introduisant la probabilité ($p = 1/N$) de chaque message possible (un nombre et son inverse ont des logarithmes opposés).

Or, la grandeur la plus utilisée en thermodynamique est précisément l'entropie, qui, pour un système quelconque, a pour valeur

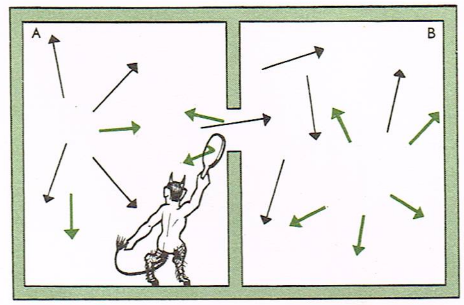
$$\mathbf{S = K \log P}$$

(K, constante de Boltzmann; **P**, probabilité de l'état considéré du système).

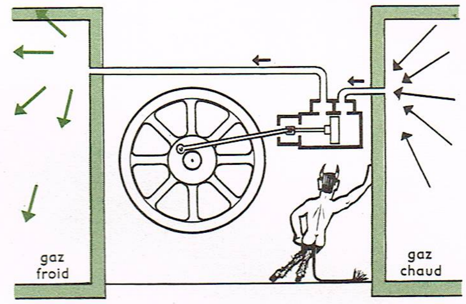
Le second principe de la thermodynamique exprime simplement que les systèmes matériels évoluent vers l'état le plus probable

$$\Delta S \geqslant 0.$$

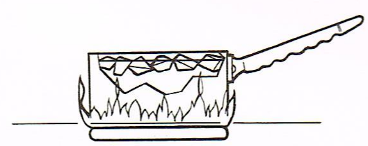
Aussi, dans l'expérience du démon de Maxwell, l'état où les molécules sont également réparties dans chacune des



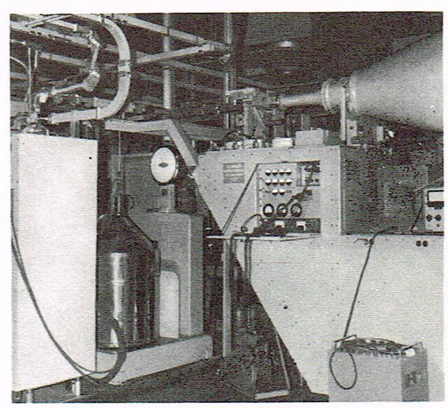
Le démon de Maxwell en rébellion contre les lois de la thermodynamique?



Le démon se retire et cède la place à une machine thermique qui fournit du travail; le second principe est violé.

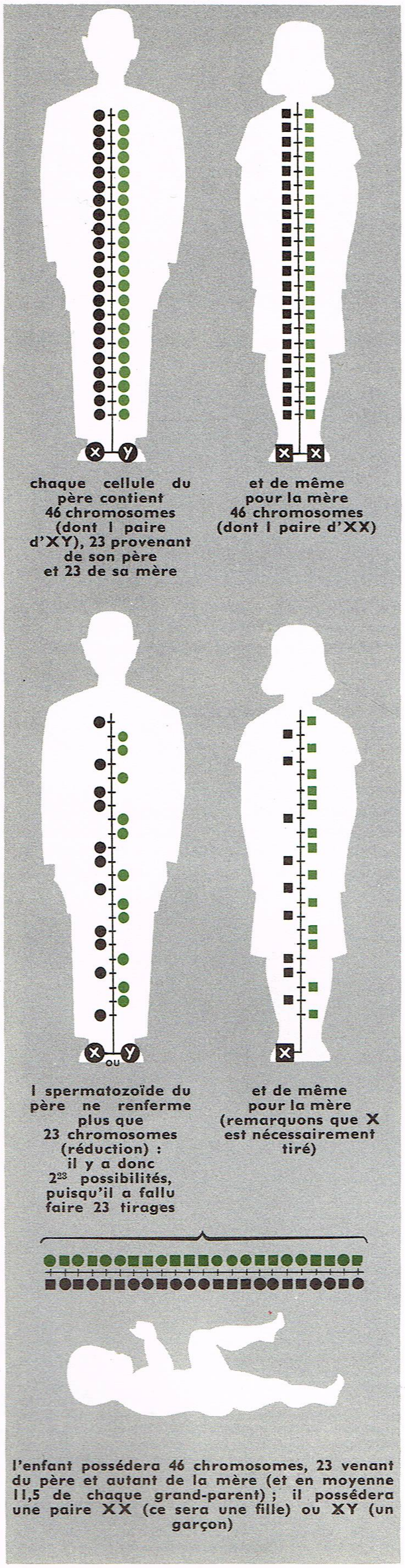


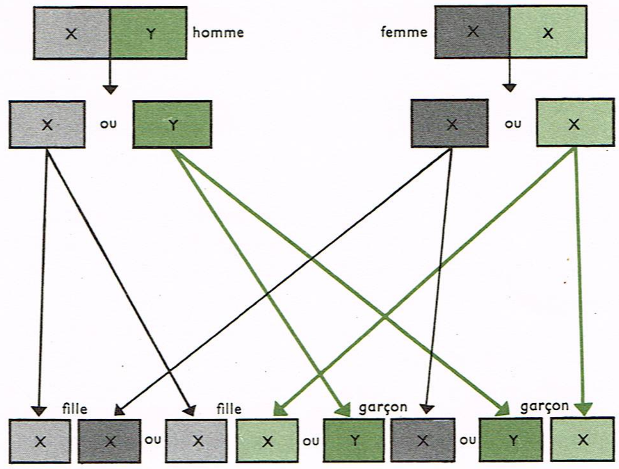
Un paradoxe tout aussi grand : la flamme se réchauffe pendant que l'eau de la casserole se prend en glace; en fait, ce phénomène est possible, mais infiniment peu probable.



Ce qui est vrai des gaz l'est évidemment des électrons qui s'agitent librement dans un métal. Mais, là, le physicien se fût volontiers passé de ce phénomène, qui est néfaste; c'est que, si ténues que soient les fluctuations de ces mouvements autour de leur valeur moyenne, les extraordinaires moyens d'amplification qui sont mis en œuvre pour la réception des signaux (T. V., T. S. F., etc.) ne font malheureusement pas la distinction entre ces signaux et cette agitation et les traitent avec le même énorme coefficient multiplicateur. Résultat, on entend cette agitation. C'est le bruit de fond, inévitable : vraie bête noire de tous les ingénieurs de l'électronique, et dont nous avons vu le rôle néfaste qu'il joue en empêchant une transmission sans perte de l'information. Une solution, refroidir le gaz. Ce que l'on fait dans le radome de Pleumeur-Bodou.

Phot. R. Goarts (doc. Air liquide).





Mathématique et hérédité : détermination du sexe (v. p. 265).

pièces est infiniment plus probable (loi des grandes séries, que nous avons déjà bien analysée) que celui où toutes les molécules chaudes sont groupées dans une pièce. Les théoriciens ont pu parler de la « mort thermique » qui attendait — tôt ou tard — l'univers consommant son entropie au prix de l'égalisation des températures de tous les corps. Dans cet état final, aucune énergie ne pourrait voir le jour et aucune action ne serait plus possible. Précisons que cette réjouissante situation, même si l'on adopte des vues aussi pessimistes et aussi discutables, n'est pas pour demain.

L'information qu'utilise précisément le démon de Maxwell pour retourner le sens d'évolution de phénomènes paraît être en ce sens l'antidote de l'entropie. Bien plus, la forme mathématique des deux grandeurs présente une analogie, pour ne pas dire une identité frappante :

$$\mathbf{S = K \log P \,;\, I = -\log_2 p}$$

(P, p sont tous deux des probabilités).

pourquoi vivons-nous du passé au futur ?

Cela ne pouvait manquer d'attirer l'attention des savants et d'intéresser leur sagacité. Léon Brillouin en a déduit un principe tout à fait remarquable, qui consiste, pour tenir compte du rôle de l'information, en réaction contre l'accroissement de désordre prédit par la thermodynamique, à remplacer la vieille et célèbre inégalité

$$\Delta S \geqslant O$$

par la suivante :

$$\Delta S - \Delta I = \Delta (S - I) \geqslant 0.$$

L'information croît moins vite que l'entropie (tout gain en information est nécessairement payé par une variation plus grande de l'entropie) :

$$\mathbf{\Delta I \leqslant \Delta S.}$$

Cette relation, qui conduit L. Brillouin à appeler négentropie l'information, « exorcise » déjà d'une façon satisfaisante le démon de Maxwell; la diminution d'entropie correspondant au rangement des particules chaudes dans une pièce est compensée par la consommation d'information nécessaire pour que le diable manœuvre correctement sa minuscule porte.

Mais l'application de cette relation, dont on commence à apercevoir la richesse, est beaucoup plus vaste. Elle explique notamment certains paradoxes auxquels le phénomène « vie » donne lieu.

L'ensemble du système « maçon + tas de briques » n'évolue-t-il pas — si du moins l'ouvrier n'est pas totalement idiot et s'il a appris à faire des murs — vers le système beaucoup moins probable « maçon + mur »? L'entropie a diminué, mais au prix d'une consommation d'énergie. Faut-il préciser qu'un être vivant consomme beaucoup d'information (elle lui est aussi nécessaire que le pain, et, du reste, le pain est d'abord organisation bien particulière et improbable de molécules, donc essentiellement support d'information)?

Le principe de Léon Brillouin est donc toujours vérifié : on n'a pu découvrir d'expérience où le gain d'information fût supérieur à la perte d'entropie que nécessitent sa création et sa communication.

Phot. U. S. I. S. et A. F. P.

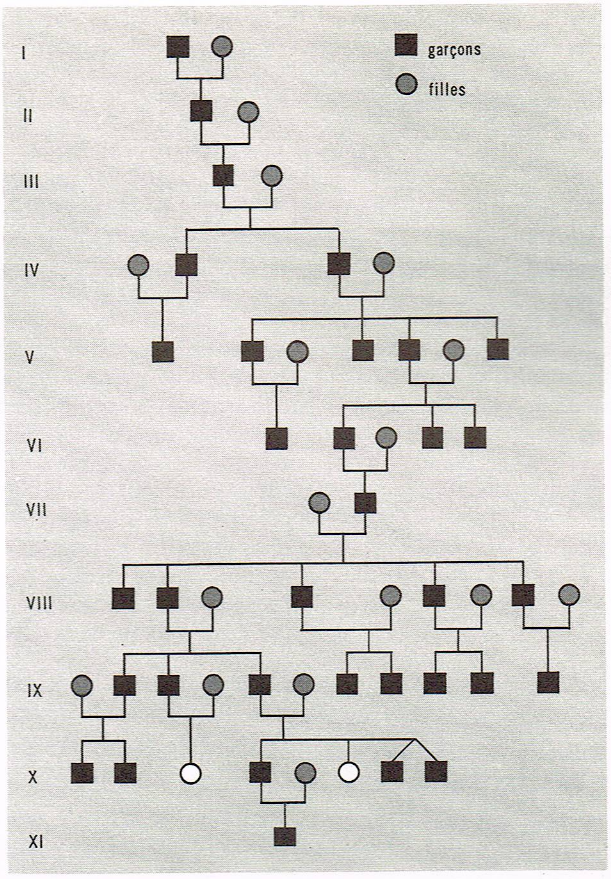
C'est précisément cette nécessité de gagner de l'information, donc de gagner de la négentropie, qui oblige, ainsi que n'ont pas tardé à le réaliser les théoriciens, l'homme à vivre du passé au futur.

la biologie et les lois de l'hérédité

Mais l'analyse bioprobabiliste ne se borne pas à étudier l'évolution des êtres vivants. On a déjà parlé à deux reprises de l'information génétique portée par l'A. D. N.; il est temps de parler d'un problème assez voisin : celui qui est posé par les lois de l'hérédité.

Le spermatozoïde porte, comme on sait, 23 chromosomes. Chacun d'eux est formé d'une série de gènes (particules ne dépassant pas 0,1 μ), supports fondamentaux des caractères de l'individu. De même pour l'ovule. Si bien que l'œuf porte après fécondation 46 chromosomes, avec apports égaux de chacun des parents. Et toutes les cellules du corps de l'enfant contiendront ces mêmes 46 chromosomes. La remarque banale qu'il ressemble davantage à son père ou à sa mère n'exprime qu'une apparence; l'égalité de l'hérédité est absolue. Cependant, une question se pose. Si toutes les cellules du corps possèdent 46 chromosomes, comment les cellules reproductrices n'en ont-elles que 23? Par l'intervention d'un phénomène capital, appelé « réduction chromosomique » : de chaque paire de chromosomes, un seul se retrouve dans l'ovule ou le spermatozoïde. La multitude des causes qui président à ce choix fait qu'aucune combinaison n'a plus de chances d'être obtenue que l'autre. *C'est le hasard qui commande l'hérédité.* Chaque spermatozoïde résulte ainsi d'un jeu « à pile ou face » effectué 23 fois de suite, soit 2^{23} (plus de 8 millions de combinaisons possibles).

Remarquer l'analogie absolument fondamentale avec la sortie des rouges et noirs de la roulette. Une chance sur 8 millions correspond au cas où un spermatozoïde ne contient aucun gène du père (ou de la mère) de celui qui le porte. Un enfant a ainsi une chance sur 8 millions de n'avoir absolument rien de commun avec un de ses grands-parents. Si faible soit-elle, cette probabilité est à l'échelle humaine, et si 100 millions d'enfants naissent chaque année, le fait peut se produire douze fois.



La détermination du sexe. Arbre généalogique d'une famille où prédominent curieusement les garçons; deux filles (en blanc sur le schéma) n'apparaissent qu'à la dixième génération. L'une d'entre elles est morte en bas âge, l'autre était intersexuée et n'a pu avoir de descendance. (D'après Harris.)

Notons que si chaque ovule et chaque spermatozoïde peuvent revêtir indépendamment 8 millions de formes différentes, l'enfant d'un couple fait partie d'un ensemble de plus de 60 milliers de milliards de possibilités. Ne nous étonnons pas que les enfants d'une famille ne se ressemblent pas!

Nous voyons ainsi que nos enfants sont biologiquement la moitié de nous-mêmes et, comme dit Jean Rostand, que nous survivons effectivement en eux, puisque, sauf accident très rare (mutation), l'enfant ne possède aucun chromosome qui ne provienne de son père et de sa mère.

Mais, par le jeu des réductions chromosomiques, notre petit-fils nous représentera en moyenne à 25 p. 100 (une fois même sur 8 millions à 50 p. 100, et une fois sur 8 millions à zéro pour cent).

la détermination du sexe

Les biologistes ont découvert que, parmi les 46 chromosomes de la femme, une paire était constituée de chromosomes identiques nommés XX, et que, par conséquent, l'ovule contenait le chromosome X.

Chez l'homme, en revanche, cette paire est constituée de deux chromosomes différents, dits XY (1). Il produit donc des spermatozoïdes X ou Y à égalité de chances. La rencontre de X avec l'ovule donne la paire XX (fille), et celle d'Y la paire XY (garçon) [v. *fig.* p. ci-contre]. La probabilité de chaque cas est évidemment 1/2, mais il semble, sans qu'on sache pourquoi, que la probabilité de rencontre XY soit de 55 p. 100 contre 45 p. 100 pour XX. Il naîtrait alors 120 garçons pour 100 filles.

Mais la mortalité après naissance ramène cette proportion à 106 contre 100, tandis que la vie moyenne plus courte de l'homme rétablit l'équilibre et évite l'écrasement de la population féminine par les hommes.

(1) A moins qu'Y n'existe pas, et, dans ce cas, l'homme n'a plus que 45 chromosomes.

Il est évident que savoir influer efficacement sur la production de spermatozoïdes X ou Y permettrait de choisir le sexe de l'enfant. Cette possibilité n'est aucunement exclue à l'heure actuelle; chacun en mesurera les immenses conséquences.

Le public est souvent étonné par des familles de cinq filles ou de cinq garçons. La probabilité en est de $(1/2)^5$, soit 3 chances sur 100, chiffre non négligeable. La distribution des sexes dans les grandes familles obéit bien aux lois du hasard. (Il semble toutefois que les longues séries de garçons ou de filles dépassent un peu les prévisions. Il paraît y avoir là une anomalie intéressante.)

une physique fondamentalement probabiliste

Les schémas probabilistes dominent donc directement beaucoup de problèmes, mais il faut aller plus loin encore.

Toute information est limitée, disions-nous. Le voudrions-nous à tout prix, serions-nous assez déraisonnables pour sacrifier toute notre existence à la seule détermination infiniment précise d'un seul des paramètres de l'univers (par exemple le nombre π ou l'épaisseur du livre que vous êtes en train de lire), même cet espoir est fallacieux.

La définition d'un objet (les deux bords d'une table ne sont jamais infiniment parallèles), la précision de sa mesure, la connaissance que nous en avons sont limitées. A plus forte raison, la somme de nos connaissances peut être très grande, et croître sans cesse (vivre, c'est s'informer), elle ne sera jamais infinie, ne fût-ce que parce que aucun canal physique ne peut transporter une information infinie.

Il nous faut donc agir en fonction de tout ce que nous connaissons et compléter la citation précédente : « Vivre c'est s'informer », ce qui nous permettra une action plus raisonnée et plus efficace. Disons encore : « Vivre, c'est prévoir; prévoir, c'est pourvoir. »

La probabilité est la mesure de notre connaissance (et non de notre ignorance, comme il a été dit méchamment).

Dans tous les exemples donnés jusqu'à présent, il est clair que le calcul des probabilités est le seul outil pratique à notre disposition (faute de temps et de moyens), et qu'aucune cause essentielle ne s'oppose à la connaissance certaine des phénomènes : on sonde l'opinion pour éviter la consultation générale, on joue à la boule parce qu'on ne peut résoudre en un instant la foule de problèmes de physique exacte qui se posent à son lancement, on ignore le sexe de son prochain enfant par la difficulté extrême de connaître les innombrables facteurs chimiques et mécaniques qui le déterminent. On ne connaît que le mouvement moyen d'une molécule d'oxygène, car il est difficile de la suivre séparément.

La physique moderne ajoute : non seulement difficile, *mais absolument et fondamentalement impossible*, et pour deux raisons.

La première est que raisonner sur une molécule déterminée d'oxygène n'a aucun sens, car les molécules sont identiques. *Pouvoir nommer un objet implique de savoir le distinguer des autres.* Qu'en est-il des objets identiques? On peut les discerner momentanément par la position qu'ils occupent dans l'espace! Qu'ils se cachent un instant, et la confusion devient totale. Deux jumeaux parfaits entrent dans une pièce. L'un précède l'autre. Je peux l'appeler A et l'autre B. L'un d'eux ressort. Lequel? Aucun moyen de le savoir. Bien plus, la question n'a pas de sens, puisque aucun d'eux ne se distingue de l'autre, comme deux molécules d'oxygène, comme deux boules de billard. Ainsi, le physicien considérera des électrons, des atomes — et non pas tel électron, tel atome —, non par difficulté ou par ignorance, mais par une conséquence essentielle de l'identité des particules.

La deuxième est que **faire de la physique c'est observer, et observer c'est perturber.** On ne voit jamais que l'effet sur un atome, une molécule, un électron, un neutron de notre instrument le plus fin : la lumière.

Cette perturbation inévitable, véritablement inhérente au fait scientifique, a été traduite par la loi de Heisenberg.

Le produit de l'incertitude sur la position d'un objet par l'incertitude sur sa quantité de mouvement est supérieur à la constante h de Planck.

Cette constante est précisément celle qui intervient dans les phénomènes d'émission de la lumière. Tout cela est extrêmement logique et solide, mais entache irrémédiablement le phénomène physique élémentaire de l'impossibilité de le connaître exactement. Son évolution est seulement probable, et la science, probabiliste par essence.

Ainsi on parvient à des concepts qui choquent d'abord l'esprit, mais apparaissent bien vite comme les seuls raisonnables ; on ne dira plus qu'un photon de lumière se propage dans l'espace, puisqu'on ne peut pas bien le localiser : le physicien sait seulement qu'il a davantage de chances de se trouver ici que là. Les équations de propagation vont tout naturellement s'appliquer non à la position — incertaine — du photon, mais à la loi de probabilité de cette position : on dit que les ondes de la mécanique ondulatoire sont des « ondes de probabilité ». Cela veut dire que les lois de mouvement s'appliquent avant tout aux chances de trouver le photon. Sir A. S. Eddington, savant britannique remarquable, propose à ce propos d'excellentes images ; l'incertitude que nous avons sur la position d'un électron, d'un photon, est semblable à un brouillard qui remplirait une pièce, dont la densité en un point serait proportionnelle aux chances qu'on aurait de trouver la particule en ce point. Si nous ne savons absolument rien, le brouillard est uniforme. Mais nous avons fait diverses expériences, à la suite desquelles nous pensons que l'électron est probablement au centre, au milieu de la pièce, où le brouillard est plus épais. Il devient le « fog » où se cache l'électron. Bien plus, les mouvements du brouillard figureront les mouvements de l'électron, et l'on sait que l'air peut vibrer en ondes sonores.

Telles sont les ondes de probabilité, qu'il est maintenant temps d'étudier de plus près.

Nous voici maintenant parvenus au seuil des grandes synthèses. Jusqu'ici, les chemins que nous avons parcourus nous sont relativement familiers : la mécanique, l'optique, l'électromagnétisme nous satisfont et nous rassurent. Même la relativité restreinte, qui couronne le bel édifice ondulatoire de la physique classique, ne nous est pas tellement étrangère. Nous nous habituons à la notion d'espace-temps, d'invariant relativiste, d'identité entre la masse et l'énergie : les collégiens s'y font très bien et la relativité passe merveilleusement dans les grandes écoles. Enfin la physique des particules est infiniment agréable. Elle admet de nombreuses images à notre échelle. Nous représentons volontiers le proton, l'électron comme de petits grains et, lorsqu'ils sont rapides, nous en suivons l'évolution comme on suit celle de billes de billard ou de boules de pétanque en espérant des collisions. Rien n'est plus parlant que les beaux clichés de transmutations nucléaires dans les chambres à bulles, dont les grands enfants reconnaissent fort bien les phénomènes.

Et pourtant toutes ces constructions agréables avec des modèles à notre échelle — exemples des billes pour les noyaux, des planètes pour les électrons des atomes, de la houle pour la lumière —, toutes ces constructions sont loin d'être satisfaisantes. Quand on quitte la dimension de notre bonne mécanique, quand on va vers le très petit, on pressent bien qu'il n'est plus possible de conserver les notions familières. La lumière n'est pas une onde, puisqu'on peut l'émettre comme une succession de projectiles de mitraillette : l'effet photo-électrique le prouve bien. Elle n'est pas non plus corpusculaire, puisqu'on voit les belles figures de diffraction, inexplicables par des projectiles. Elle doit comporter les deux aspects à la fois, et cela n'est vraiment pas imaginable pour notre esprit classique.

Tout craque d'ailleurs si l'on cherche une représentation satisfaisante. Les orbites des électrons planétaires ne sont pas toutes valables, nous affirme-t-on ; seules certaines sont possibles : celles dont le moment cinétique est multiple entier de la très petite constante de Planck. Le succès de Bohr est d'avoir trouvé les règles du jeu, qui sont incompréhensibles ; bien plus, l'électron satellite décrit son orbite indéfiniment sans émettre de rayonnement, ce qui est contraire à toutes les conceptions de l'électromagnétisme classique. On peut multiplier les exemples : comment expliquer l'émission radio-active alors que les particules α, que l'on voit sortir, ne devraient pas pouvoir quitter le noyau, n'ayant pas l'énergie suffisante pour en franchir les barrières? Bien plus, comment expliquer toute la chimie, l'association des atomes en molécules? Peut-on comprendre quelque chose à cette liaison chimique si simple et si abstraite en même temps, cette petite barre que les chimistes placent entre H et H, entre deux atomes d'hydrogène, pour constater qu'ils peuvent s'unir et former une molécule ?

Oui, il nous faut tout reprendre. Notre mécanique est en effet impuissante à l'échelle atomique ; les notions simples de point matériel, de vecteur force, auxquelles nous sommes si habitués et sur lesquelles elle est fondée, ne sont plus valables quand on les transpose à d'aussi petites dimensions. La notion d'un parfait déterminisme, décrivant l'évolution d'un système

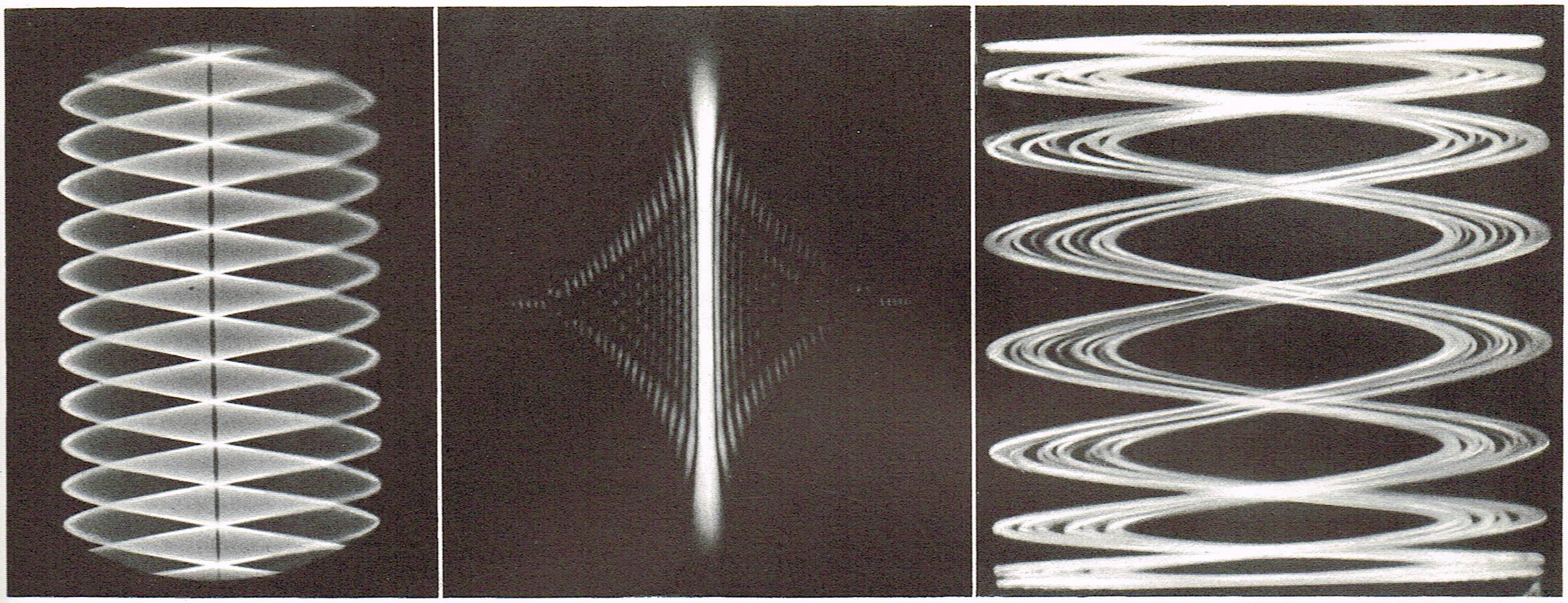
physique à partir des positions et vitesses des points connus à un certain moment de façon précise, cette notion doit être modifiée.

Ainsi une nouvelle mécanique devait naître, plus générale et plus raffinée que l'ancienne, pour expliquer l'atome de façon valable, sans contradictions, tout en se raccordant à la mécanique classique lorsque les phénomènes sont à notre échelle. Ce fut un immense effort, aboutissant depuis 1924, grâce au génie de Louis de Broglie, puis de Schrödinger, de Heisenberg, de Pauli, de Dirac, etc., à une admirable synthèse : la mécanique ondulatoire ou quantique. Il faut bien voir ceci : c'est que, pour englober les phénomènes atomiques dans une explication cohérente, il a fallu abandonner des notions familières, qui ne sont en fait que des approximations, ainsi la notion de connaissance précise de la position ou de la vitesse ; on se tourne vers une vue du monde matériel où l'affirmation de présence et d'absence s'estompe, où la "probabilité de présence" règne seule, où s'introduit une incertitude fondamentale que les célèbres relations d'Heisenberg expriment avec une admirable simplicité.

Mais il n'est pas facile de changer notre représentation du monde, de nous adapter allégrement à des concepts nouveaux, étrangers à nos modes de pensée, exprimés par un formalisme mathématique difficile et inhabituel. C'est dur, même pour les physiciens. Alors peut-on expliquer cette révolution, au moins la faire pressentir honnêtement au lecteur non physicien, parfois non scientifique, de cet ouvrage ? C'est ce que nous avons tenté dans les derniers chapitres de ce tome, et nous avons fait pour cela un important effort d'adaptation. Les spécialistes des chapitres "la matière : ondes ou corpuscules?", "mécanique quantique" et "architecture de la matière", jeunes physiciens dominant parfaitement leur discipline, ont passé et repassé leur texte au crible de critiques exigeantes, simplifiant et clarifiant au maximum la formulation des nouvelles conceptions.

On ne peut aller beaucoup plus loin sans tout défigurer, sans anémier les motivations, sans amollir à l'excès la belle architecture de la mécanique quantique, en un mot sans tricher. Mais le lecteur ne doit pas s'y tromper : même simplifiées au maximum, même introduites avec ménagement et patience, les pages qui suivent ne se lisent pas comme celles d'un journal à grand tirage. Elles demandent au lecteur un effort d'attention, de concentration, une continuité dans cet effort. Mais les trésors intellectuels qu'il pressentira lorsque, parvenu à la fin de l'ouvrage, il aura pu embrasser l'ensemble des conceptions et des explications lui apparaîtront comme assez remarquables pour justifier largement son effort.

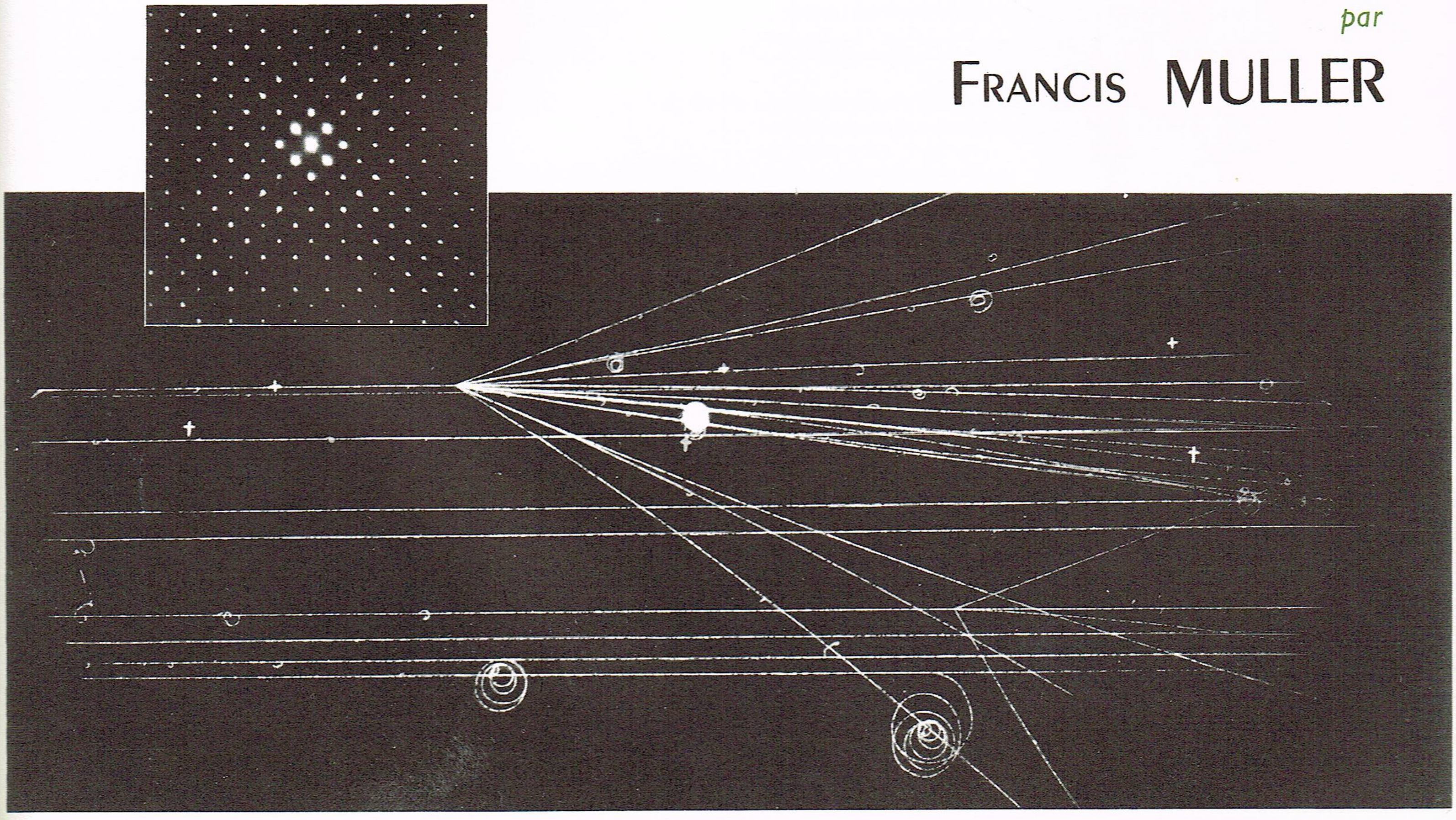
Louis LEPRINCE-RINGUET



la matière
ONDES OU CORPUSCULES

par

FRANCIS MULLER



Doc. C.S.F. et C.E.R.N.



Niels Bohr.



Louis de Broglie.



Erwin Schrödinger.



Werner Heisenberg.

Pourquoi donc une nouvelle mécanique, alors que la mécanique classique et sa forme relativiste permettent de résoudre, avec une précision remarquable, des problèmes très variés?

Tel Le Verrier, les astronomes peuvent prédire la position exacte de Neptune à n'importe quel moment; les balisticiens sont capables d'amener une fusée à rencontrer Vénus à un rendez-vous fixé dans un an et à 200 millions de kilomètres de la Terre; plus modestement, l'efficacité de l'artillerie de marine est proverbiale. Enfin, lorsque les vitesses sont telles que la relativité est nécessaire, les exemples abondent où calculs et expériences sont en accord. La ronde fantastique des électrons ou des protons dans les grands accélérateurs n'en est-elle pas la meilleure preuve?

Dans tous ces cas, la connaissance des forces qui s'exercent et les données de la position et de la vitesse initiale du projectile suffisent à déterminer entièrement la trajectoire.

Si les fusées et même les obus de marine manquent parfois leur objectif, c'est à cause d'une connaissance expérimentale imparfaite de certaines données (poussée de la fusée, freinage dans l'air et vitesse initiale de l'obus) et non d'une insuffisance fondamentale de la théorie.

Lorsque le système est composé d'un grand nombre de particules, les complications mathématiques sont telles que l'on doit renoncer à connaître le mouvement de chaque particule; mais les méthodes statistiques permettent de traiter l'évolution de l'ensemble du système. C'est ainsi que la statistique classique, due à Boltzmann, donne une interprétation des lois fondamentales de la thermodynamique et explique la plupart des propriétés des gaz, par exemple.

Bien que certains de nos exemples soient très récents, les lois qui les concernent étaient connues et admises à la fin du XIXe siècle. La théorie de la relativité ne date que de 1905, mais, en dépit des âpres controverses auxquelles elle donna lieu, elle s'inscrit dans la même ligne de pensée.

L'ensemble de ces lois et de leurs conséquences constitue la physique classique, dont le caractère fondamental est, on le voit, un déterminisme rigoureux : l'évolution d'un système peut être, en principe, prédite dans ses moindres détails, si l'on connaît complètement son état à un certain moment.

L'habitude aidant, cette proposition était passée du rang de simple conséquence des lois de la physique à celui de vérité première.

Le mot « déterminisme » tel qu'il est défini ici n'a pas nécessairement une implication philosophique.

En effet, à la question « Le monde de l'an 2000 serait-il donc déjà déterminé? » le physicien classique aurait pu répondre :

« En principe oui, si tous les phénomènes relèvent des quelques lois fondamentales que je connais. Mais, comme il est humainement impossible de connaître tous les paramètres définissant le monde actuel, il est non moins impossible de prédire le monde de l'an 2000. D'ailleurs, ne suis-je pas déjà incapable de décrire complètement l'évolution d'un millimètre cube de gaz? »

Pourtant, ce n'est pas le fait d'être satisfaisante pour l'esprit qui prouve la validité d'une théorie physique, mais bien son accord avec l'ensemble des faits expérimentaux qui en relèvent. Ainsi, toute la beauté formelle des équations de Maxwell (« Est-ce un dieu qui a écrit ces lignes? », disait Boltzmann) et la splendide logique de la théorie de la relativité ne seraient que pures constructions de l'esprit si elles n'étaient confirmées par maintes expériences, au premier rang desquelles se placent la production et la détection des ondes électromagnétiques par Hertz et la célèbre expérience de Michelson (toutes deux en 1887) sur l'invariance de la vitesse de la lumière.

De même, le succès des lois de la mécanique classique ou relativiste quand on les applique au mouvement des planètes ou à l'accélération des particules paraît bien prouver la validité de ces lois. Et pourtant, nous allons le voir, elles sont incapables d'expliquer les propriétés des atomes. Comme Niels Bohr aimait à le dire, une vérité profonde s'oppose à une autre vérité profonde. De cette confrontation, allons-nous voir naître la Vérité ?

Phot. U.S.I.S., Erwing Marton, N.Y.T.

une contradiction majeure : les raies lumineuses des atomes

Au moment même de l'apogée de la physique classique, les patientes recherches des spectroscopistes montraient que la lumière émise par un élément donné ne contenait que certaines fréquences, bien définies, toujours les mêmes : le spectre lumineux d'un atome est un spectre de raies, aussi caractéristique de ce type d'atome que les empreintes digitales le sont d'un individu.

Or, en 1911, Rutherford découvrit que presque toute la masse de l'atome — dont les dimensions sont d'environ 1 angström (10^{-8} cm) — est concentrée dans un noyau, de charge positive et de dimensions extrêmement réduites (inférieures à 10^{-12} centimètre). Attractions électrostatique et gravitationnelle obéissant à la même loi (en r^{-2}), les électrons décriraient des ellipses autour du noyau; l'analogie avec le système solaire paraît complète **(modèle planétaire de Rutherford, 1911)**.

Mais une difficulté se présente tout de suite : au cours de leur ronde, les électrons rayonneraient de la lumière, donc perdraient de l'énergie, car émettre un rayonnement, c'est perdre de l'énergie, et ils « spiraleraient » vers le noyau avec une vitesse angulaire de plus en plus grande, en émettant donc une lumière de fréquence de plus en plus élevée. Cela est en contradiction avec les fréquences bien définies observées dans les spectres.

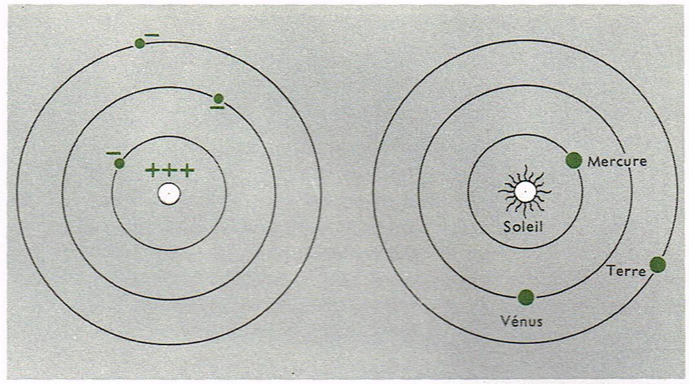
Une question se pose donc : pourquoi, alors que les lois de la mécanique permettent n'importe quelle orbite elliptique initiale et que les lois de Maxwell prévoient l'émission d'un spectre continu, observe-t-on seulement certaines fréquences bien définies? Comment rendre compte de ces nombres, associés à chaque atome comme un numéro d'inventaire?

Le Suisse Balmer fut un précurseur dans cette voie : il trouva (1885) une relation mathématique donnant les neuf fréquences alors connues dans le spectre de l'hydrogène.

Ecrite sous forme moderne, voici la formule, étonnamment simple, qui donne la valeur de la fréquence ν_n d'une des raies; si l'on donne à n des valeurs entières, à savoir 3, 4, 5..., on obtient les valeurs exactes des diverses raies :

$$\nu_n = R_H c \left(\frac{1}{2^2} - \frac{1}{n^2}\right).$$

c est la vitesse de la lumière, R_H est une constante, dite « constante de Rydberg pour l'hydrogène », et pour laquelle les mesures actuelles les plus précises donnent $R_H = 109\,677{,}58$ C. G. S. A l'époque de Balmer, les mesures spectroscopiques étaient



Comparer l'atome de lithium dans le modèle planétaire de Rutherford et le système solaire, dont les trois premières planètes seules ont été représentées.

déjà très précises, et il fallait pousser jusqu'à la cinquième décimale pour trouver une possibilité de désaccord entre la longueur mesurée et la longueur d'onde calculée par la formule de Balmer. Un accord si remarquable peut difficilement être dû au hasard!

Depuis, d'autres raies de l'hydrogène ont été découvertes — en fait une infinité de raies, qui peuvent se déduire de la formule plus générale $\nu = Rc\left(\frac{1}{m^2} - \frac{1}{n^2}\right)$. m et n sont alors deux nombres entiers : si l'on prend $m = 2$, on retrouve les raies de Balmer. Des formules analogues, quoique légèrement plus complexes, dues à Rydberg, s'appliquent aux métaux alcalins (sodium, potassium, argent...), et Ritz put énoncer une sorte de règle pratique pour la détermination des raies, appelée « principe de combinaison ». D'après Ritz, à chaque atome correspondent des séries de « termes » :

$$\nu'_1, \nu'_2, \ldots, \nu'_n \ldots; \nu''_1, \nu''_2, \ldots, \nu''_m \ldots;$$

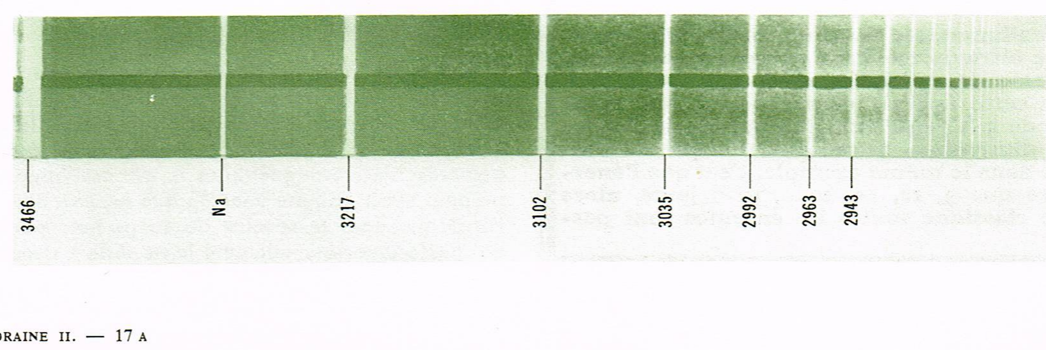
et toutes les fréquences que l'atome excité peut émettre s'obtiennent en combinant ces termes :

$$\nu = \nu''_m - \nu'_n.$$

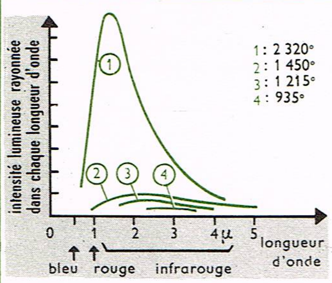
Ainsi un ensemble en apparence incohérent de nombres compliqués (les fréquences) peut se déduire d'un tableau de nombres entiers.

Le principe de Ritz est la clé qui permit à Bohr d'interpréter le mystère des spectres atomiques et à Heisenberg de fonder une nouvelle mécanique.

Spectre de la lumière émise par le potassium à l'état de vapeur. Chaque raie lumineuse a une longueur d'onde, c'est-à-dire une couleur bien précise, et peut se caractériser par un nombre (les spectroscopistes utilisent en général le nombre de longueurs d'onde par centimètre).







plus le métal est chaud dans le convertisseur, plus blanche est la lumière qu'il rayonne.

spectre du corps noir

LE CORPS NOIR. Le vocable mystérieux de « corps noir » désigne un corps idéal qui émet autant de lumière qu'il en absorbe. Une enceinte fermée isolée de l'extérieur (un four) est un très bon exemple. On conçoit facilement qu'une telle enceinte atteigne rapidement un état d'équilibre qui ne dépend que de la température des parois. La répartition de l'intensité lumineuse entre les différentes longueurs d'onde est ce qu'on appelle le spectre du corps noir.

La figure ci-dessus représente le spectre d'un corps noir à différentes températures. Plus la température est élevée, plus la longueur d'onde moyenne est basse. La lumière devient de moins en moins rouge, de plus en plus blanche ; phénomène bien connu et utilisé pour apprécier la température d'un bain de métal en fusion.

On peut calculer ce spectre en supposant que les organes émetteurs-récepteurs de lumière sont des électrons de la paroi, oscillant autour de leur position d'équilibre (chacun constitue un « oscillateur harmonique ») et en expliquant les phénomènes d'émission et d'absorption par la théorie de Maxwell. Appliquant alors la statistique classique de Boltzmann, qui donne le nombre d'oscillateurs en fonction de l'énergie, on trouve le spectre de Rayleigh, courbe physiquement inadmissible, puisque l'intensité lumineuse devient de plus en plus grande lorsque la longueur d'onde diminue : c'est la « catastrophe ultraviolette ».

En utilisant le même modèle physique, mais en introduisant l'hypothèse *ad hoc* que l'énergie d'un oscillateur de fréquence ν doit être un multiple entier du produit de ν par une certaine constante h, Planck obtint de façon purement mathématique un spectre reproduisant parfaitement le spectre expérimental. Cet accord remarquable est réalisé pour une valeur extrêmement petite de h :

$$h = 6{,}6 \quad 10^{-27} \text{ C.G.S.}$$

Ainsi l'énergie fondamentale $h\nu$ d'un oscillateur émettant de la lumière ultraviolette est d'environ $5 \cdot 10^{-19}$ joule. En fait, la petitesse de ce chiffre n'est pas réellement surprenante, car cette énergie correspond à une élongation maximale de 5 Å ($5 \cdot 10^{-8}$ cm), ce qui est de l'ordre de grandeur des dimensions atomiques. Ce qui est surprenant, toujours dans le même exemple, c'est que l'énergie ne puisse être que 5, 10, 15, 20,..., 10^{-19} joule, alors qu'en mécanique classique toutes les énergies sont possibles.

les quanta de lumière

Pour comprendre comment Bohr a pu donner une interprétation physique du « principe de combinaison » de Ritz, simple jeu arithmétique de nombres, il faut revenir sur le problème de la nature de la lumière.

Sans remonter aux Hébreux, qui se contentèrent du vœu que la lumière soit, ni aux Grecs, pour qui le phénomène de la vue était dû à la rencontre de particules émises par l'objet et de particules émises par l'œil — première apparition, un peu compliquée il est vrai, de la théorie corpusculaire de la lumière —, ni même à Newton, à la suite duquel on crut pendant un siècle que les rayons lumineux n'étaient que des trajectoires de corpuscules (les grains de lumière), il faut rappeler que, depuis Young et Fresnel (1819), la nature ondulatoire de la lumière était universellement admise. Mieux, Maxwell avait donné, en 1865, de la vibration lumineuse, une interprétation physique simple, d'après laquelle, par exemple, les rayons lumineux que la Terre reçoit du Soleil ne seraient qu'une partie d'une gigantesque onde électromagnétique continûment émise par le Soleil.

l'effet photo-électrique

En 1887, Hertz démontrait brillamment la réalité des ondes électromagnétiques. Mais, par une curieuse ironie du sort, il découvrait dans la même expérience l'effet photo-électrique, c'est-à-dire la propriété qu'a la lumière visible d'arracher des électrons à une surface métallique. Cela n'a rien de surprenant si la lumière est due à la propagation d'un champ électromagnétique : le champ électrique fait vibrer les électrons du métal, et l'on conçoit que cette vibration puisse être assez forte pour que les électrons s'en échappent. Mais l'aspect quantitatif du phénomène n'est pas explicable par la seule théorie de Maxwell : dans cette théorie, augmenter l'intensité de la lumière, c'est augmenter le champ électrique, et l'on s'attend dans ces conditions à voir augmenter l'énergie des électrons libérés. Or, il n'en est rien : si l'on augmente l'intensité de la lumière, c'est le nombre des électrons libérés qui augmente; leur énergie reste rigoureusement la même. En revanche, l'énergie des électrons libérés croît proportionnellement à la fréquence de la lumière incidente.

Pour expliquer ces deux phénomènes, il faut admettre, avec Einstein, qu'un faisceau lumineux est constitué non par la propagation d'une seule onde, mais par une sorte de grêle de petits paquets de lumière, et que chacun de ces paquets a une énergie proportionnelle à la fréquence. C'est le quantum d'énergie lumineuse : $\mathbf{W = h\nu}$.

On revient ainsi à la conception corpusculaire de Newton. Mais les phénomènes d'interférence nous obligent cependant à admettre que la lumière présente aussi un caractère ondulatoire. **Ces deux aspects paraissent absolument contradictoires.** Ainsi le quantum d'énergie introduit par Einstein eut du mal à faire reconnaître son existence, tant la théorie ondulatoire était admise comme vérité; **et le mieux est d'admettre que ce sont deux aspects complémentaires de la même réalité.**

le corps noir

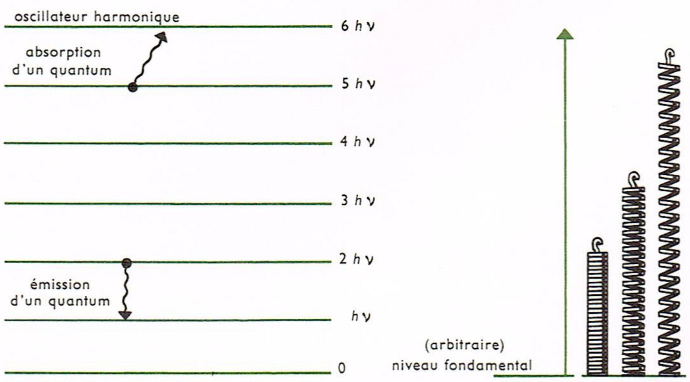
Il faut noter ici que, quelques années avant l'introduction du quantum de lumière par Einstein, Max Planck avait, pour expliquer le spectre du corps noir, postulé la quantification des énergies des « oscillateurs », émetteurs et récepteurs des rayonnements électromagnétiques : un oscillateur de fréquence ν ne peut avoir comme énergie que $h\nu$, $2h\nu$, $3h\nu$, ..., $nh\nu$. h est une constante, dont le spectre du corps noir permet la mesure et qui porte tout naturellement le nom de « constante de Planck ».

Doc. Sallac, phot. Jolin.

Mais la foi en la théorie de Maxwell était si bien ancrée dans les esprits que Planck n'alla pas jusqu'à la conséquence logique de son hypothèse : le morcellement du rayonnement électromagnétique lui-même en petits paquets d'énergie bien définie $h\nu$.

La découverte directe du quantum de lumière $h\nu$ — par l'effet photo-électrique — permet d'interpréter l'hypothèse de Planck sur la quantification de l'énergie de l'oscillateur harmonique : l'oscillateur a certains niveaux d'énergie possibles, multiples entiers de $h\nu$, et il « saute » de l'un à l'autre par émission ou absorption d'une quantité bien définie d'énergie lumineuse $h\nu$.

Ces phénomènes peuvent être représentés sur un diagramme d'énergie :



Chacun des « états » possibles de l'électron oscillant est représenté par un niveau dont l'ordonnée représente l'énergie du système. L'oscillateur passe d'un état à l'état voisin en émettant ou en absorbant de la lumière. De même, pour tendre un ressort, il faut lui fournir une certaine énergie — qu'il peut restituer en se détendant. D'après Planck, cette énergie est nécessairement un multiple du quantum $h\nu$.

En conclusion, tout ce passe comme si l'oscillateur émettait ou absorbait une particule d'énergie $h\nu$ C'est cette « particule » que l'on appelle **grain de lumière (ou photon),** ce qui ne veut pas dire qu'il faille se représenter ce photon comme un grain rond et infiniment petit.

le modèle atomique de Bohr

La voie est maintenant tracée pour l'interprétation des spectres atomiques, si mystérieux. Les « termes » spectraux de Ritz, dont la différence donne la fréquence des raies lumineuses, ne rappellent-ils pas les niveaux d'énergie successifs de l'oscillateur, dont la différence donne le quantum d'énergie lumineuse? Effectivement il suffit de multiplier la relation de Ritz

ν_{nm}	=	ν_n	—	ν_m
fréquence émise		terme spectral « n »		terme spectral « m »

par la constante de Planck, pour trouver une équation susceptible d'interprétation physique :

$h\nu_{nm}$	=	$h\nu_n$	—	$h\nu_m$.
		énergie de l'atome		
énergie du photon émis		avant émission		après émission

Cela signifie que l'atome, comme l'oscillateur harmonique, n'a que certains niveaux d'énergie possibles. Le même phénomène se retrouve d'ailleurs dans les spectres des molécules (spectres de bandes), et paraît bien être une caractéristique des systèmes composés d'un petit nombre de particules élémentaires, alors que rien d'analogue ne semble exister à notre échelle. Cela est certainement en rapport avec la petitesse de la constante h; lorsque les énergies sont très grandes, la différence quantifiée $h\nu$ entre deux niveaux voisins apparaît comme infiniment petite : expérimentalement, le spectre des énergies possibles apparaît continu. **C'est pourquoi il n'y a pas quantification apparente du mouvement d'un pendule,** qui est pourtant un oscillateur harmonique à notre échelle.

Il n'y a donc pas nécessairement contradiction entre notre expérience courante et les faits atomiques.

Reste à expliquer ces derniers.

La théorie quantique de Bohr fut la première tentative faite en ce sens. En deux mots, Bohr admet que les lois de la mécanique classique s'appliquent au mouvement des électrons, mais que seules certaines orbites privilégiées sont possibles et que, sur ces orbites, l'électron tourne sans émettre de rayonnement. Ces orbites sont définies par des conditions « quantiques », c'est-à-dire faisant intervenir des nombres entiers (c'est ainsi que la rotation de l'électron est quantifiée : son moment cinétique ne peut prendre comme valeurs que les multiples entiers de h). Si le succès pratique de la théorie de Bohr a été considérable, grâce en particulier aux travaux de Sommerfeld, il convient de noter que Bohr lui-même ne la considérait que comme une première approximation.

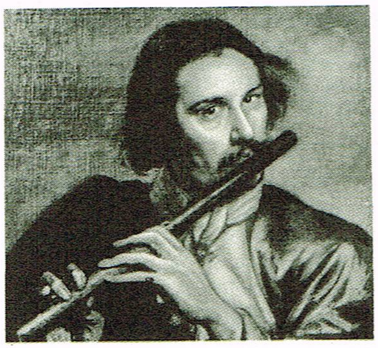
D'abord les conditions quantiques apparaissent un peu comme « parachutées » pour les besoins de la cause, alors que l'on aimerait les voir découler de principes plus généraux. Cette remarque d'ordre philosophique ne saurait suffire à condamner la théorie de Bohr. Mais, aussi, elle est incapable de justifier le fait que l'électron ne rayonne pas sur son orbite, ni d'expliquer la transition d'une orbite à une autre. Enfin, certains détails très fins des spectres — qui étaient inconnus à l'époque — lui échappent complètement.

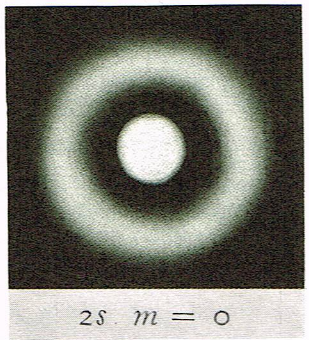
On peut dire que la théorie de Bohr (1913) était l'ultime tentative faite pour expliquer les phénomènes sans détruire le cadre de la mécanique classique. Ses insuffisances, ainsi que l'existence d'autres phénomènes inexpliqués — telle la valeur anormale de la chaleur spécifique des corps à basse température — suffirent à convaincre de la nécessité d'une autre mécanique, tout au moins à l'échelle des particules. Un peu comme la relativité est résultée de la réflexion profonde d'Einstein sur les concepts de temps et d'espace, cette mécanique va se déduire d'une révision du concept de particule. Ainsi de Broglie, Schrödinger et Heisenberg introduisent, vers 1925, une représentation des phénomènes élémentaires toute différente et, à sa suite, plus qu'un nouveau langage : une nouvelle façon de penser.

Comme a dit plus tard un célèbre physicien : « En 1925, il n'y eut pas seulement un grand progrès en physique théorique; ce fut comme si, en quelques mois, notre vie avait changé. »

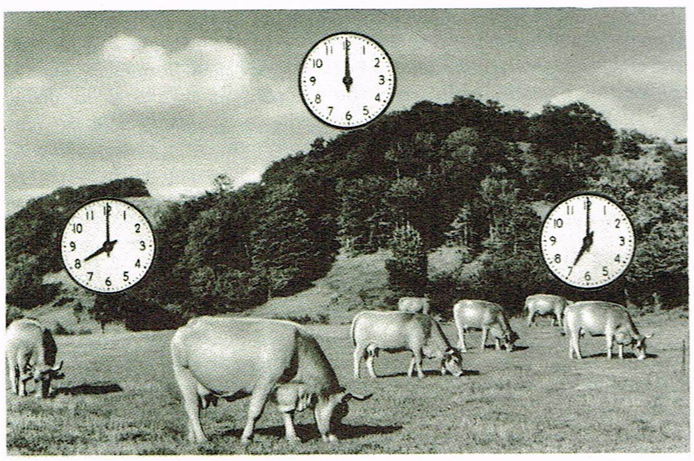
vers une nouvelle mécanique :

Louis de Broglie postule la nature ondulatoire de la matière

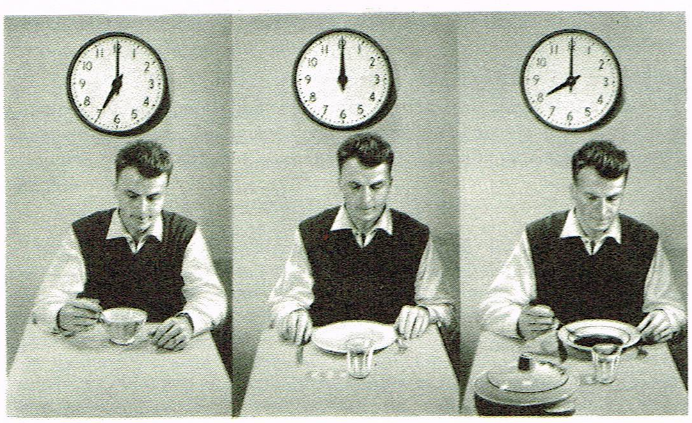




La flûte n'émet que certaines notes, donc certaines fréquences pures. L'atome n'a que certaines énergies, donc certains états propres. Les deux systèmes ont un point commun : ils sont quantifiés.



La vache mange de façon continue...



... l'homme de façon discontinue : c'est une « suite discrète » de repas.

C'est Louis de Broglie qui eut le premier l'idée de représenter le mouvement d'une particule par la propagation d'une onde (1923). Nous ne reproduirons pas ici son raisonnement, mais nous essaierons de montrer comment l'introduction d'une onde pour décrire un mouvement est susceptible d'expliquer le phénomène de quantification.

analogie entre les mouvements des électrons atomiques et les résonances des instruments de musique

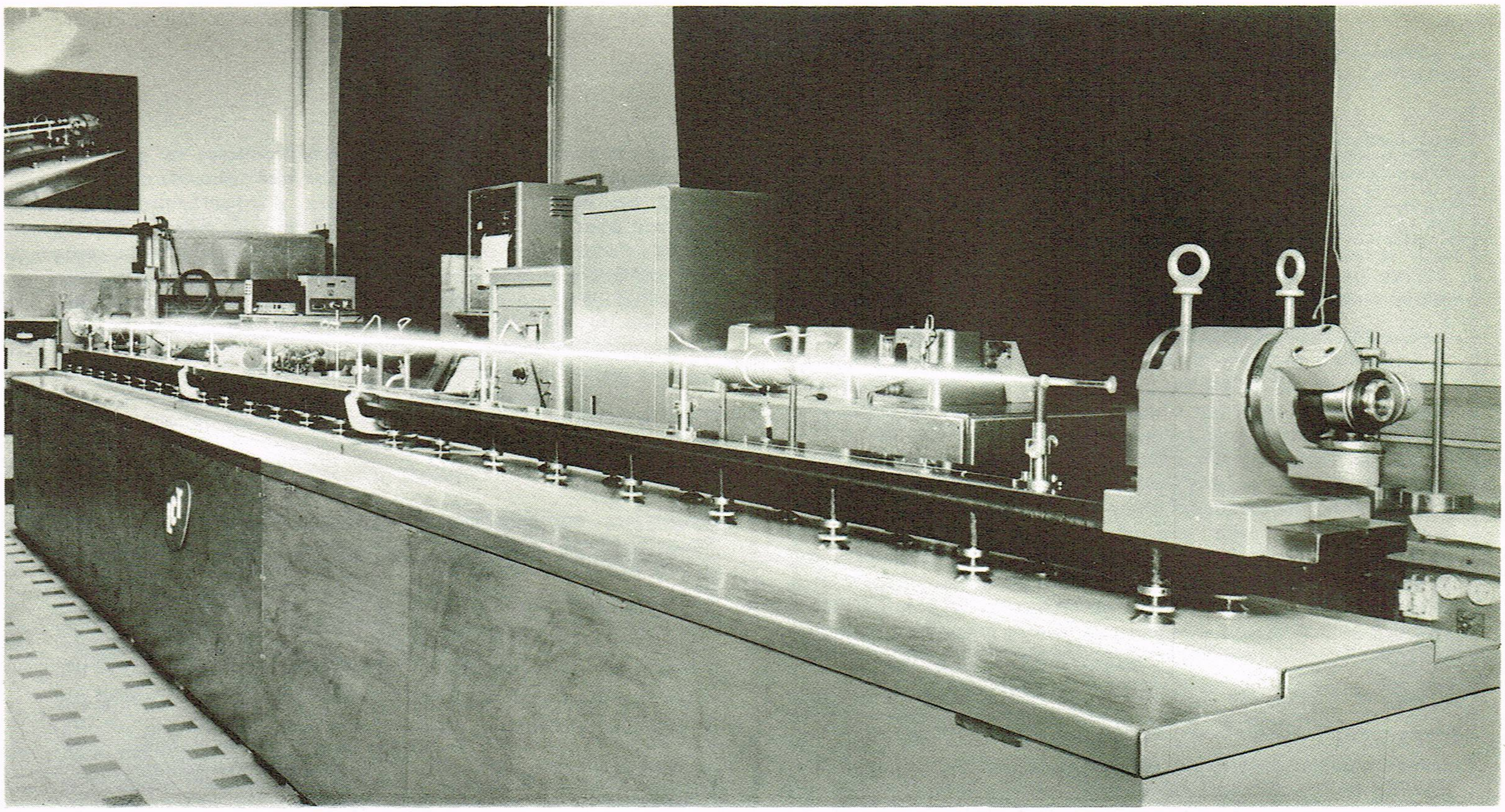
On a vu par ailleurs qu'une corde de violon ou un tuyau d'orgue ne peuvent émettre que certains sons, composés du fondamental et de ses harmoniques, par exemple la_1 (fréquence $f_0 = 435$ cycles/seconde), la_2 (fréquence $2f_0 = 870$ cycles/seconde), la_3 ($3\ f_0$), etc. Pour prendre un exemple un peu plus compliqué, une membrane circulaire tendue sur un cadre rigide (un tambour) ne peut, elle aussi, vibrer que suivant certaines fréquences ν_0, ν_1, ..., ν_n, que l'on peut déduire d'une équation, dite « équation de Bessel », laquelle traduit mathématiquement le problème.

Un cas similaire est présenté par les cavités, telles que celles des klystrons, magnétrons, etc., utilisées dans le domaine des hyperfréquences en radio-électricité, où seules des ondes électromagnétiques de fréquences données peuvent subsister en régime stationnaire; ici encore — tout au moins si les cavités ont une forme simple —, ces fréquences peuvent aisément se calculer à partir de l'équation de propagation des ondes électromagnétiques et des conditions « aux limites » sur les parois de la cavité. L'expression « conditions aux limites » signifie qu'il faut tenir compte de ce qui se passe aux extrémités de la partie vibrante. Ainsi, pour une flûte, le son ne sera pas le même si l'on bouche ou si l'on ouvre l'extrémité du tuyau; pour un tambour, les modes de vibration dépendent de la forme du tambour, etc. Ce sont ces « conditions aux limites » que l'on doit introduire dans les équations, en posant par exemple que la vibration est nulle à l'emplacement du bord rigide du tambour.

Dans tous ces exemples, le fait qu'une vibration (sonore ou électromagnétique) soit contenue dans une certaine région entraîne l'existence d'une suite discrète de modes de vibrations possibles. Le terme de **« suite discrète »** que nous aurons à utiliser souvent signifie que, parmi un ensemble continu de valeurs *a priori* possibles d'une grandeur physique, seules existent certaines de celles-ci, séparées par des intervalles finis; leur nombre peut être infini, mais à chacune d'elles on peut attacher un nombre entier (ensemble dénombrable). Par exemple, la vache mange de façon pratiquement continue, alors que les moments de la vie où l'homme prend ses repas successifs constituent une suite discrète.

Revenons maintenant à l'atome. L'étude du spectre de raies montre que l'énergie de l'électron et, par conséquent, les éléments de son orbite ne sont pas susceptibles de prendre n'importe quelle valeur. L'analogie avec les systèmes vibratoires est frappante.

Phot. Despoisse, Larousse.



Le laser produit une onde lumineuse qui est l'illustration la plus parfaite d'une onde plane monochromatique.

la dualité onde-corpuscule

Cela incite à penser que la description du mouvement de l'électron par la propagation d'une onde pourra permettre d'expliquer l'existence d'états privilégiés, parmi tous ceux qui sont possibles en théorie classique. Telle est l'« hypothèse ondulatoire » : à toute particule est associée une onde; nous préciserons plus loin comment, et quelle est la signification de cette onde.

lumière et matière

D'ailleurs la lumière ne présente-t-elle pas, suivant les expériences, des caractéristiques tantôt ondulatoires (phénomènes d'interférence), tantôt corpusculaires (effet photo-électrique); pourquoi n'en serait-il pas de même de la matière? Ainsi, un faisceau d'électrons, qui, jusqu'à la formulation de l'idée de De Broglie — et sa confirmation par l'expérience —, devait toujours apparaître comme un jet de grains chargés extrêmement petits et insécables, pourrait dans certains cas se comporter comme une onde — c'est-à-dire, par exemple, donner lieu à des figures d'interférence.

Remarquons toutefois que, bien que cette analogie des propriétés de l'électron et du photon satisfasse sans doute notre goût pour la symétrie, cette symétrie entre la lumière et la matière est loin d'être totale. Ainsi, elle n'existe pas dans un domaine important : la lumière peut être à volonté détruite ou créée — le photon disparaît dès qu'il a cédé son énergie, dans l'effet photo-électrique par exemple —, alors que, sauf circonstances exceptionnelles, électrons et nucléons sont indestructibles : ils peuvent être plus ou moins énergiques, mais ils existent *ad aeternam*.

L'hypothèse de génie de Louis de Broglie, formulée en 1923 dans sa célèbre thèse, demandait donc confirmation. La première fut apportée par la théorie elle-même, puisque c'est en partant de la description ondulatoire de la matière que Schrödinger bâtit la mécanique quantique et put, dès 1925, l'appliquer avec succès à l'explication de plusieurs phénomènes. La confirmation expérimentale directe, c'est-à-dire la production de figures d'interférence par un faisceau d'électrons, attendit quatre ans (1927). Avant de voir comment, il nous faut expliciter un peu l'hypothèse ondulatoire.

un peu de théorie : la longueur d'onde de De Broglie

Dans la théorie ondulatoire de Huygens et de Fresnel, la lumière peut se représenter par la propagation d'une vibration (le champ électrique de la théorie de Maxwell, mais nous n'avons pas besoin ici de spécifier la nature de cette vibration, que nous appellerons simplement u).

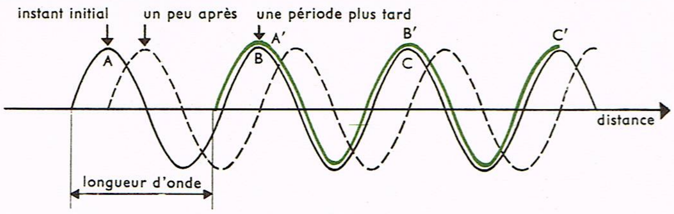
Une lumière de couleur pure (par exemple, la lumière jaune émise par un laser) est dite « monochromatique ». Cela signifie que cette lumière ne contient qu'une seule des fréquences que renferme l'arc-en-ciel. La propagation de cette lumière dans un milieu transparent se représente alors par deux nombres : sa fréquence ν, c'est-à-dire le nombre d'oscillations par seconde ou « périodes T » de la vibration lumineuse, et sa longueur d'onde λ, c'est-à-dire la distance que parcourt la lumière en une période. Si v est la vitesse de propagation de la vibration lumineuse, on a évidemment la relation :

$$\lambda = v\mathrm{T} = v/\nu.$$

Dans le vide, la vitesse v a pour valeur :

$$c = 300\,000 \ \text{km/s}.$$

Un cas particulièrement simple est constitué par la propagation d'un faisceau lumineux parallèle, c'est-à-dire provenant d'une



source très lointaine, ou encore d'une lampe placée au foyer d'une lentille. L'onde qui décrit un tel faisceau est dite *onde plane*.

La courbe représentant la grandeur de la vibration lumineuse à un certain moment le long d'un des rayons du faisceau est une *sinusoïde*; elle n'est pas sans rappeler les vagues de la mer, d'où le nom d'onde (en allemand et en anglais, « ondes » se traduit par « vagues » : *Wellen, waves*). Un peu plus tard, la sinusoïde s'est déplacée, et exactement une période plus tard, la crête qui était en A se substitue à la crête qui était en B : la vibration s'est propagée d'une longueur d'onde. Remarquons qu'une période T plus tard, la sinusoïde recouvre exactement la sinusoïde initiale : si l'on photographie la surface de la mer à deux instants séparés par la période de la houle, les vagues ont le même aspect, comme si rien n'avait changé. Cette remarque, justifiée pour la houle, aura des conséquences fondamentales pour la représentation d'une particule en mouvement.

Passons maintenant à la description corpusculaire de la lumière. Quel est l'équivalent d'une onde plane monochromatique? C'est un faisceau parallèle, c'est-à-dire un jet continu de photons, tous de même énergie. Or, à chaque photon (V. chap. RELATIVITÉ) sont rattachées l'énergie $E = h\nu$ (ν étant la fréquence de l'onde monochromatique équivalente) et la quantité de mouvement.

$$p = \frac{E}{c} = \frac{h\nu}{c} = \frac{h}{\lambda}.$$

Ainsi cette comparaison très simple entre onde et corpuscule de lumière nous conduit à rapprocher les données qui définissent le mouvement du corpuscule, énergie E et quantité de mouvement p, des données qui définissent l'onde associée : fréquence ν et longueur d'onde λ.
Les relations précédentes se mettent sous une forme particulièrement simple et symétrique en introduisant la constante $\hbar = \frac{h}{2\pi}$ et en remplaçant la fréquence ν et la longueur d'onde λ par la pulsation ω et le nombre d'ondes réduit k (1).

Pour la lumière, les relations entre les grandeurs caractérisant les deux descriptions possibles, ondulatoire et corpusculaire, s'écrivent alors :

DESCRIPTION		RELATIONS
ONDULATOIRE	CORPUSCULAIRE	
Pulsation ω	Énergie E	$E = \hbar\omega$
Nombre d'ondes k	Quantité de mouvement p	$p = \hbar k$

L'hypothèse de Louis de Broglie est la suivante : à tout mouvement de particule (particule matérielle cette fois, et non plus seulement lumineuse) est associée une onde et les nombres ω et k caractérisant cette onde se déduisent de l'énergie E et de la quantité de mouvement p de la particule par les relations

$$p = \hbar k$$
$$E = \hbar\omega.$$

C'est un raisonnement par induction : nous savons que ces relations sont vraies pour le photon; admettons qu'elles le sont aussi pour l'électron, le proton, etc. (2).

(1) $\omega = 2\pi\nu$ et $k = \frac{2\pi}{\lambda}$. La pulsation ω et le nombre d'ondes k sont définis par : $E = h\nu$, écrit alors $E = \hbar\omega$; $p = \frac{h}{\lambda}$ devient $p = \hbar k$.

(2) Une raison fondamentale pour y croire est fournie par la relativité : E et pc forment un quadrivecteur; il en est donc de même pour ω et kc.

Or, fait remarquable, cette idée révolutionnaire est parfaitement confirmée par l'expérience. **Avec des faisceaux d'électrons on peut faire des anneaux d'interférence, tout comme avec des ondes lumineuses.** On peut même mesurer, d'après l'écartement de ces anneaux, la longueur d'onde associée au faisceau d'électrons, et constater que cette longueur d'onde est identique à la longueur prévue par de Broglie, laquelle se déduit de la relation notée ci-dessus et réécrite ainsi :

$$\lambda = \frac{h}{p}.$$

C'est la célèbre relation de De Broglie, et λ s'appelle la **longueur d'onde de De Broglie.**

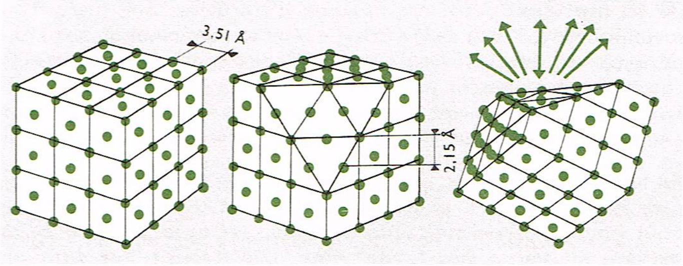
première vérification expérimentale : Davisson et Germer font interférer les électrons

Voyons un peu ce que cela donne pour des électrons de vitesse v très inférieure à la vitesse de la lumière. On peut alors écrire $p = mv = \sqrt{2mT}$, où m est la masse et T l'énergie cinétique $\frac{1}{2}mv^2$ des électrons. En général, cette énergie leur est fournie par une différence de potentiel V; elle est alors égale à eV (e, charge électrique de l'électron) et s'exprime commodément en électron-volts. La formule numérique s'écrit $\lambda = \frac{12{,}27}{\sqrt{V}}$ Å.

On voit que, pour les électrons de quelques électron-volts, la longueur d'onde est de l'ordre de l'angström (rappelons que la longueur d'onde de la lumière visible est de l'ordre de 6 000 Å). Pour réaliser des interférences, il faut donc disposer de réseaux dont le pas soit aussi de l'ordre de l'angström; de tels réseaux sont fournis par les mailles des cristaux, et la possibilité de réaliser des interférences avec des cristaux est bien connue dans le cas des rayons X (rayonnements électromagnétiques, de même nature que la lumière, mais dont la longueur d'onde est beaucoup plus petite).

L'expérience fut réalisée par hasard, en 1927, par Davisson et Germer au moyen d'un « canon à électrons » envoyant un faisceau pratiquement parallèle d'électrons, d'énergie bien définie, réglable entre 36 et 68 électron-volts, perpendiculairement à une face d'un gros cristal de nickel.

En observant la variation de l'intensité diffusée en fonction de l'angle, ils constatèrent un maximum à 0°, puis un minimum, puis un maximum secondaire à environ 50° (pour des électrons de 54 électron-volts). Or, avec un réseau de pas d (ici égal à 2,15 Å), on s'attend à un tel maximum lorsque est



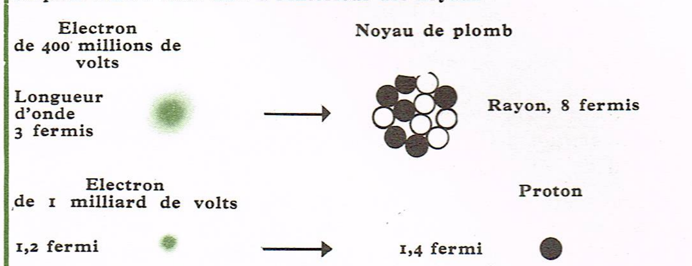
L'expérience capitale de Davisson et Germer.
A gauche, **structure cubique d'un cristal de nickel;** *au centre*, **coupe de cristal montrant la facette utilisée par Davisson et Germer pour la diffusion des électrons;** *à droite*, **un pinceau d'électrons arrive perpendiculairement à cette facette; les électrons sont diffusés par le réseau que constitue le cristal et on constate que cette diffusion est particulièrement forte suivant certaines directions.**

Une conséquence de la nature ondulatoire de l'électron et des autres particules élémentaires.

Classiquement, on peut s'imaginer l'électron comme une sorte de granule extrêmement petit. Dans tous les calculs classiques, on peut d'ailleurs le représenter par un point. En mécanique ondulatoire, il faut considérer l'électron comme un « objet », dont la « dimension » est de l'ordre de la longueur d'onde de De Broglie. Il apparaît donc d'autant plus petit qu'il est plus énergique. Ainsi, un électron accéléré sous une tension de 1 milliard de volts se présente avec une dimension d'environ 1 milliardième de micron, unité connue sous le nom de fermi, en mémoire du grand physicien italien.

L'électron est alors « plus petit » que les noyaux (rayon : 1,4 fermi pour l'hydrogène, 8 fermis pour le plomb).

En bombardant une cible avec un faisceau d'électrons très énergiques, on peut donc « viser fin » à l'intérieur des noyaux :



C'est ainsi que l'on a exploré, à 100 millions de volts et plus, la répartition des charges électriques à l'intérieur des noyaux, et, à 1 milliard de volts, l'intérieur du proton lui-même. Le même phénomène se retrouve lorsque l'on bombarde les noyaux avec des protons; plus le proton est énergique, plus il vise avec finesse à l'intérieur du noyau. C'est pourquoi l'on construit des accélérateurs de plus en plus grands.

satisfaite la relation de Bragg. Puisque $\lambda = d \sin \theta$, on aura ici : $\lambda = 2{,}15 \sin 50^0 = 1{,}65$ Å, alors que la longueur d'onde de De Broglie est $\lambda = \frac{12{,}27}{\sqrt{54}} = 1{,}67$ Å.

L'excellent accord se retrouve dans d'autres expériences analogues et prouve la validité de la relation $\lambda = h/p$. Cette relation est si bien admise de nos jours qu'on utilise la diffraction par les cristaux en sens inverse : connaissant l'énergie des électrons, on en déduit le pas d du réseau. C'est un puissant moyen d'exploration de la structure des cristaux.

il n'y a pas que les électrons

Mieux encore : le même type d'expérience est réalisé avec des faisceaux moléculaires — des jets de molécules d'hydrogène ou d'atomes d'hélium, dont on observe la diffraction par la surface d'un cristal. En recueillant dans un récipient les particules diffractées à un angle θ, on constate la formation, dans le récipient, d'hydrogène ou d'hélium — avec une pression qui est maximale pour la valeur de θ calculée d'après les relations de Bragg et de Louis de Broglie : $d \sin \theta = \lambda = h/p$.

Encore plus extraordinaire : le même phénomène est observé avec des faisceaux de neutrons!

Sans doute l'hypothèse ondulatoire est-elle vraie pour le grain de sable. Mais alors, du fait de la petitesse de h, la longueur d'onde de De Broglie est énormément plus petite que la dimension du grain, et les effets ondulatoires deviennent inexistants; **c'est pourquoi la mécanique quantique, nécessaire à l'échelle microscopique, est inutile à notre échelle, ainsi qu'il a déjà été remarqué.**

quand faut-il utiliser la mécanique ondulatoire?

En conclusion, nous venons de voir un type de phénomène — la diffraction de particules par des réseaux cristallins — que la théorie ondulatoire permet d'expliquer, alors que la mécanique classique en est incapable. L'analogie avec l'optique nous permet de prévoir que ce sera le cas général lorsque les particules seront confrontées avec des obstacles d'une dimension à peu près égale à leur longueur d'onde ou, plus généralement, lorsqu'elles pénétreront dans des champs de force variant beaucoup sur une longueur d'onde. Pour parler un langage plus simple, un objet en mouvement passant à travers une ouverture ne s'aperçoit guère de l'existence des parois si l'ouverture est beaucoup plus large que lui (une souris dans une porte), mais s'en aperçoit si l'ouverture est à peu près de la même dimension (trou de souris); ce que prévoit l'hypothèse ondulatoire, c'est que les dimensions des obstacles doivent se comparer à la longueur d'onde de la particule en mouvement et non à sa « dimension » au sens classique du terme. On peut immédiatement penser que la nature ondulatoire des électrons jouera un rôle dans la structure des atomes : en effet, la dimension des atomes est de l'ordre de l'angström (0,53 Å pour l'hydrogène) et l'énergie cinétique des électrons est de l'ordre de quelques dizaines d'électron-volts (13 eV pour l'hydrogène), donc leur longueur d'onde est de l'ordre de l'angström. **C'est parce que le champ électrique du noyau (champ que subit l'électron) varie beaucoup sur l'étendue d'une longueur d'onde qu'il nous faut utiliser la mécanique ondulatoire.**

En revanche, le domaine de validité de la mécanique classique (relativiste, s'il le faut) englobe non seulement le mouvement des objets macroscopiques (planètes, fusées, obus, boules de billard, etc.), mais encore celui des faisceaux de particules élémentaires (électrons, protons, etc.) dans des champs électriques et magnétiques faiblement variables (champs usuels fabriqués par l'homme). De même, l'optique géométrique — propagation en ligne droite, lois de la réflexion et de la réfraction — est une approximation parfaite lorsque l'on néglige les effets de diffraction dus à la dimension finie des diaphragmes; à cette approximation près, on montre facilement que la théorie ondulatoire de la lumière conduit aux lois bien connues de l'optique géométrique. On peut montrer aussi que la représentation d'un faisceau d'électrons par une onde de De Broglie conduit, pour le mouvement des électrons dans un champ suffisamment uniforme, à la même solution que la mécanique classique. Celle-ci apparaît ainsi comme une optique géométrique des particules (ce parallélisme avait été remarqué il y a plus de cent ans par Hamilton, qui avait énoncé pour la mécanique un principe de moindre action de forme identique au principe de Fermat, du moindre chemin optique).

pouvoir séparateur du microscope à électrons

Terminons ce paragraphe en mentionnant une autre application de la nature ondulatoire des corpuscules : nous avons vu que la longueur d'onde associée à un électron de 150 électron-volts est égale à 1 Å, c'est-à-dire est 5 000 fois plus petite que la longueur d'onde du violet. D'autre part, le pouvoir séparateur d'un microscope est limité — précisément par le phénomène de diffraction des ondes à travers l'ouverture du microscope — à une distance entre deux points donnée par la formule que les collégiens connaissent bien : $d = \frac{1{,}22\,\lambda}{\sin u}$.

On voit l'énorme avantage à travailler avec de faibles longueurs d'onde; c'est ce que permet le microscope électronique, où l'objet à examiner est « éclairé » par un faisceau d'électrons d'énergie donnée. Les électrons diffusés par les irrégularités de l'objet sont concentrés sur l'écran détecteur par des « lentilles » électrostatiques et magnétiques. On atteint ainsi en principe des dimensions de l'ordre du millième de micron (en fait, des impératifs techniques limitent plus sévèrement le pouvoir résolutif), ce qui permet l'exploration de la structure des grosses molécules organiques.

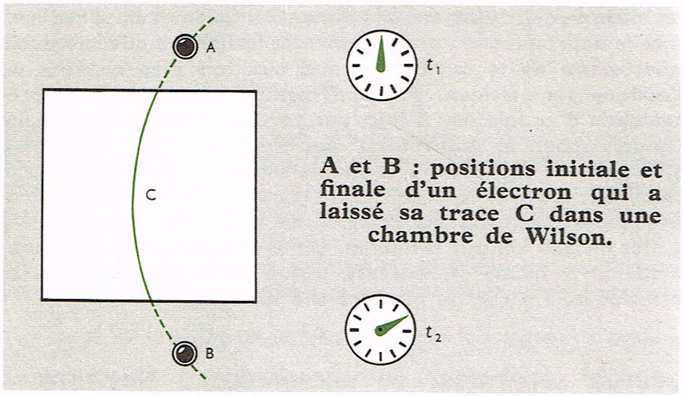
Il faut noter que, dans le microscope électronique, l'hypothèse ondulatoire n'intervient pas dans le calcul de l'optique du système, qui se fait très bien en mécanique classique (de même que, dans le microscope ordinaire, on calcule les images par l'optique géométrique), mais dans la limitation théorique apportée à la finesse de l'image d'un point par le phénomène purement ondulatoire de la diffraction.

le « paquet d'ondes »... Pourquoi ?

L'idée première de Louis de Broglie consistait à représenter par une onde plane un faisceau d'électrons d'énergie E bien définie se déplaçant en ligne droite, exactement comme nous représentons un faisceau de rayons lumineux parallèles et de couleur bien définie : $u = u_0 \cos(\omega t - kx)$, avec $\omega = \frac{E}{\hbar}$, $k = \frac{p}{\hbar}$.

Les expériences de diffraction et d'interférence des électrons, neutrons et autres particules, nous ont montré que cette association était fondamentalement correcte. Néanmoins, une telle onde n'est pas suffisante à elle seule pour représenter le mouvement de particules isolées.

En effet, ainsi qu'il a déjà été noté, une onde monochromatique se reproduit identique à elle-même au bout d'une période, et l'on ne voit pas comment on peut suivre le mouvement de la particule qu'elle est censée représenter. En fait, **où est la particule?** L'onde s'étendant sur tout l'espace, il est impossible de dire que la particule est, à un moment donné, à tel endroit plutôt qu'à tel autre. Or, malgré cela, nous savons bien qu'à un certain instant une particule a une certaine localisation. Ainsi on peut déceler les passages successifs à des temps t_1 et t_2 d'un électron dans deux compteurs à scintillation A et B, et photographier la trace C qu'il laisse dans une chambre de Wilson; on peut dire qu'à l'instant t_1 l'électron était quelque part en A, à l'instant t_2 en B, et, entre t_1 et t_2, il a décrit une courbe voisine de C.



A et B : positions initiale et finale d'un électron qui a laissé sa trace C dans une chambre de Wilson.

Plus précisément, en A la particule a perdu un peu de son énergie à l'instant t_1, et, en B, à l'instant t_2 (rappelons-nous le point de départ d'Einstein pour sa théorie de la relativité : ***pas de signal sans énergie*).**

Le mouvement d'une particule, c'est donc de l'énergie localisée qui se déplace. Or, une onde plane monochromatique ne représente pas un déplacement d'énergie. Par exemple, un bouchon flottant à la surface d'une mer animée d'une houle régulière ne se déplace pas : il monte et descend sur place.

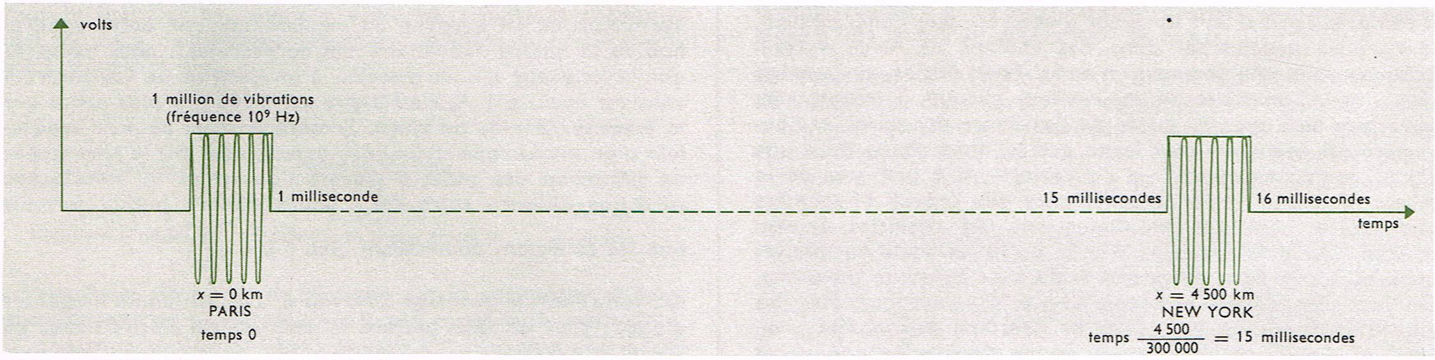


Ce qu'il faut pour représenter le mouvement d'une particule, ce n'est donc pas une onde telle que celle-ci, mais plutôt une perturbation de durée limitée et qui, elle, se déplacerait à la vitesse de la particule.

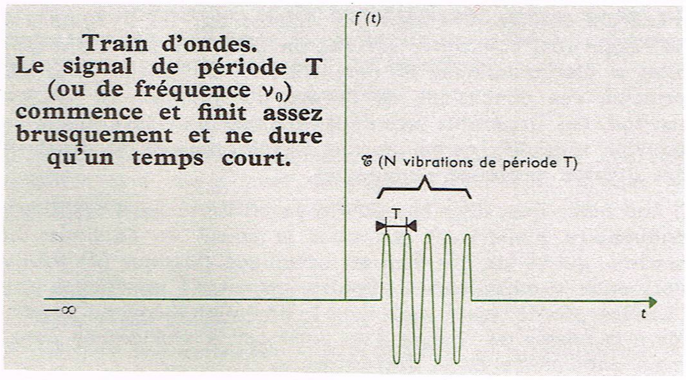
encore une analogie : la propagation d'un signal

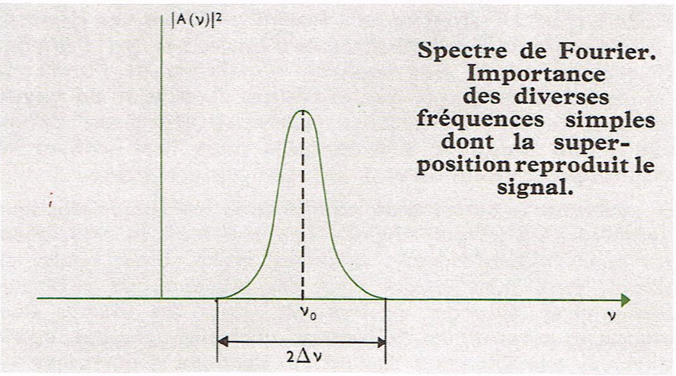
Un signal est constitué par l'émission d'une succession de vibrations électriques pendant un temps de durée limitée. C'est un « train d'ondes » et non une onde entretenue — laquelle ne donnerait dans le récepteur qu'un ronflement continu, dépourvu de toute information.

Le signal émis à Paris à 0 heure est reçu à New York à 0 heure 15 millisecondes. Il ne dure à chaque endroit qu'un temps très bref, par exemple une milliseconde. On voit l'analogie avec le passage d'une particule successivement devant deux observateurs.



Train d'ondes. Le signal de période T (ou de fréquence ν_0) commence et finit assez brusquement et ne dure qu'un temps court.





Spectre de Fourier. Importance des diverses fréquences simples dont la superposition reproduit le signal.

Microscope électronique en service à l'Institut Pasteur. *(Doc. O. D. F.)*

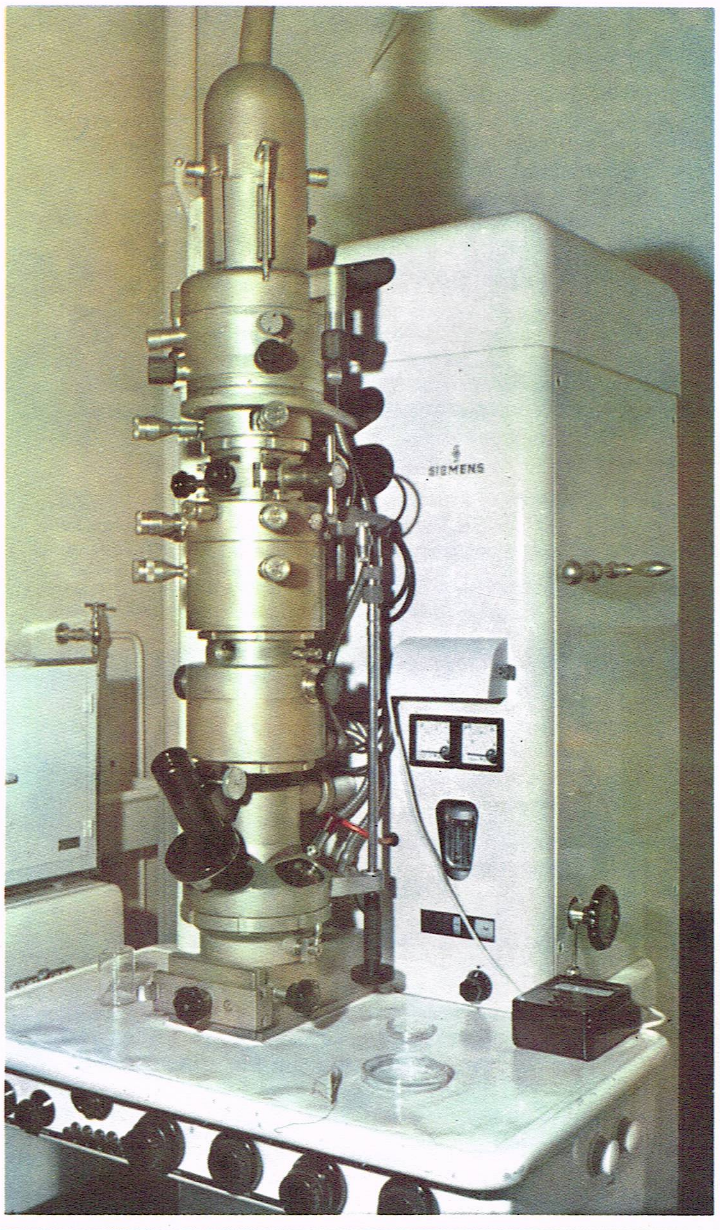
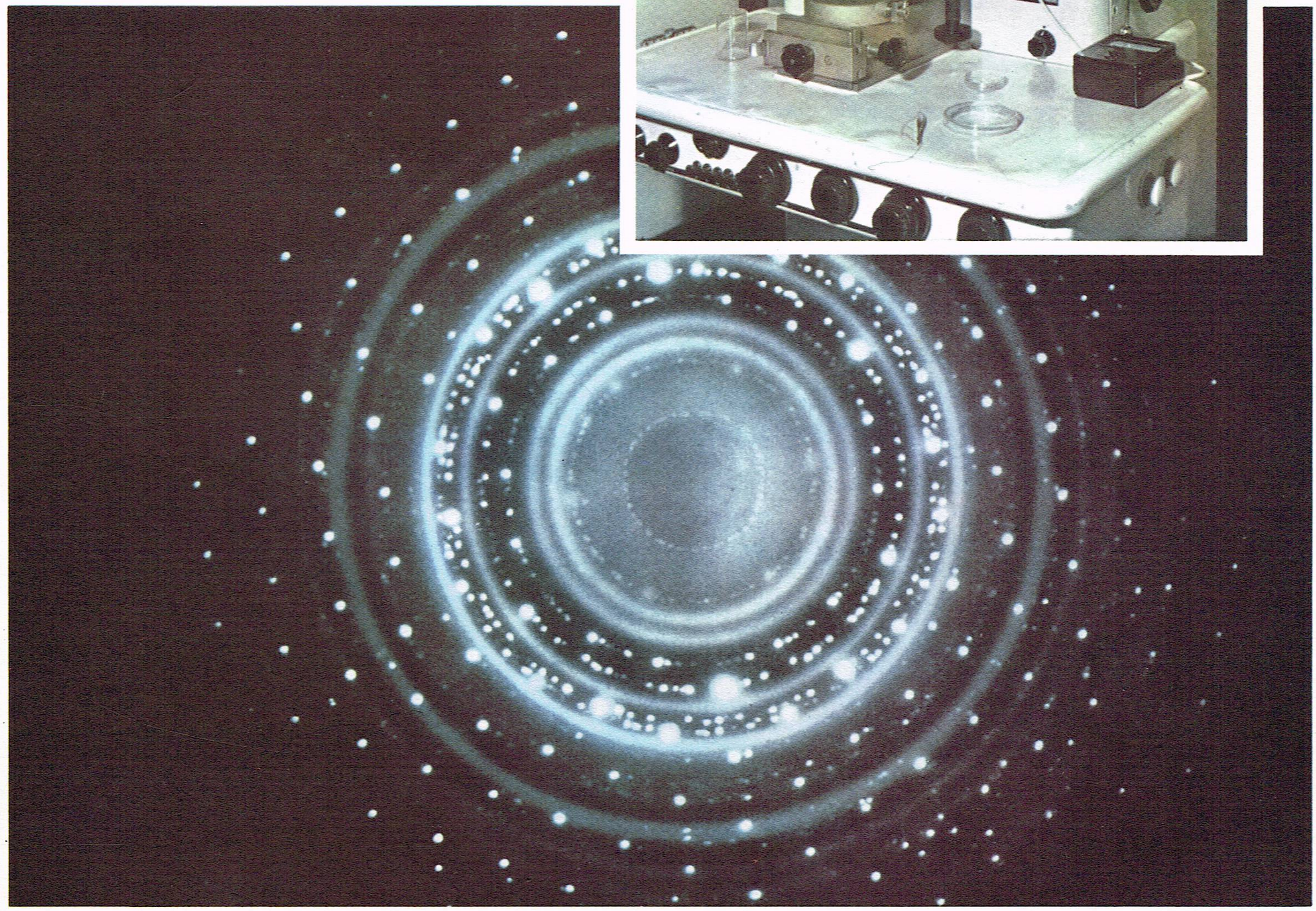


Diagramme de diffraction électronique d'une substance cristalline très finement divisée. L'intensité et le diamètre des différents anneaux permettent de calculer les intervalles entre les atomes qui constituent les cristaux, ainsi que de déterminer leur disposition géométrique. *(Cliché Zeiss.)*



Mais nous savons (v. p. 116) qu'une impulsion, si courte soit-elle et quelle que soit sa forme, peut toujours se représenter par la superposition de vibrations sinusoïdales élémentaires. Dans le cas particulier d'un train d'ondes de fréquence fondamentale ν_0 (10^9 dans notre exemple), mais durant seulement un temps τ (une milliseconde dans notre exemple), on trouve que ce train d'ondes est équivalent à une somme d'ondes sinusoïdales, chacune ayant sa propre fréquence ν et une certaine intensité $I(\nu)$.

En toute rigueur, toutes les fréquences ν doivent être considérées, pour que leur superposition reproduise exactement le train d'ondes; mais, en pratique, l'intensité $I(\nu)$ correspondant à la fréquence ν devient négligeable dès que ν s'écarte appréciablement de la fréquence ν_0. Donc, en pratique, on n'aura à composer que des fréquences très voisines.

Ici intervient une propriété mathématique, dont les conséquences vont être fondamentales : **la largeur du spectre de fréquences est inversement proportionnelle à la durée τ de l'impulsion.**

De façon plus précise, les fréquences composantes sont groupées à l'intérieur d'une bande dont la largeur $\Delta\nu$ est donnée par la relation

$$\tau \Delta \nu \sim 2\pi$$

ou encore

$$\tau \Delta \omega \sim 1.$$

En résumé, on peut représenter un nombre limité d'oscillations de fréquence ν_0 — un train d'ondes — par un paquet d'ondes (somme d'ondes sinusoïdales de fréquences ν très voisines de ν_0).

C'est cette représentation que nous allons essayer maintenant d'appliquer à une particule.

une raie lumineuse n'est jamais infiniment fine

C'est un fait d'expérience que, si l'on mesure au moyen d'un interféromètre de haute précision, du type Pérot-Fabry par exemple, la fréquence d'une raie spectrale, on ne trouve pas une valeur unique ν_0, mais un spectre de valeurs étalées entre $\nu_0 - \Delta\nu$ et $\nu_0 + \Delta\nu$ (le rapport de ν à $\Delta\nu$ caractérise la finesse de la raie; les raies naturelles ont une finesse inférieure à un million).

Ainsi, le fait d'introduire non une seule fréquence, mais un groupe de fréquences voisines, apparaît justifié pour les photons qui composent cette raie lumineuse. Nous verrons plus loin la signification profonde, pour toute particule, de l'existence d'un tel spectre.

le paquet d'ondes va à la vitesse de la particule...

L'existence non d'une seule fréquence, mais d'un groupe de fréquences voisines est essentielle pour qu'il existe un paquet d'ondes localisé dans l'espace.

Pour que notre représentation ait un sens, il est nécessaire que la vitesse de déplacement de ce paquet d'ondes, ou vitesse de groupe, v_g corresponde à celle que la mécanique classique attribue à la particule.

Effectivement, un calcul, que nous ne pouvons donner ici, montre que $v_g = \beta c$, ce qui est bien la vitesse de la particule (1).

Ainsi nous voilà devant un résultat très encourageant : nous avons cherché à représenter une particule par un groupe ou un « paquet d'ondes », et nous trouvons que ce paquet se déplace avec une vitesse qui est exactement la vitesse de la particule telle qu'on la définit habituellement. **C'est bien l'évolution du paquet d'ondes qu'il nous faut suivre pour décrire la propagation de l'énergie localisée, c'est-à-dire le mouvement de la particule.**

(1) Une seule onde a une vitesse de propagation (dite « vitesse de phase ») qui non seulement n'est pas la vitesse de la particule, mais est supérieure à celle de la lumière (pas de contradiction avec la théorie de la relativité, car, nous l'avons dit, une onde monochromatique ne transporte pas de signal).

...mais son énergie est mal déterminée

Une difficulté apparaît : le fait qu'une infinité de fréquences voisines interviennent, alors que la fréquence est, d'après la relation $E = h\nu$, strictement déterminée par l'énergie de la particule. Laquelle de toutes les fréquences voisines va-t-on choisir pour définir l'énergie de la particule? Si l'on garde la correspondance fréquence-énergie, alors l'énergie ne serait pas bien déterminée? Que peut signifier cela? Nous avons l'habitude de considérer des particules bien localisées à tel endroit, avec une énergie ou une vitesse qui peuvent être parfaitement définies si nos appareils de mesure sont parfaits. **Allons-nous voir du flou apparaître dans des notions pourtant si satisfaisantes?** Eh oui, c'est bien cela, et il nous faudra concevoir de manière tout à fait nouvelle des notions aussi « évidentes » que celles de position, vitesse, énergie.

Mais est-ce bien absolument nouveau? Non pas! Il suffit de regarder avec un peu de pénétration ce qu'est une raie lumineuse pour déjà soupçonner un problème.

l'interprétation probabiliste de Max Born

En associant à une particule un paquet d'ondes de fréquences très proches, nous avons obtenu une certaine localisation de la particule, mais on ne peut plus parler maintenant d'une fréquence unique. En fait, chaque fréquence élémentaire est représentée avec une certaine intensité $I(\nu)$ [1].

Revenons alors au problème posé précédemment. Quelle peut bien être l'énergie de la particule, sachant qu'entre énergie et fréquence existe la relation $E = h\nu$?

Prenons le cas d'une collection d'atomes excités identiques. La raie spectrale que l'on observe est due à l'enregistrement de l'ensemble des photons émis, après passage dans un système dispersif (prisme, réseau). On pourrait donc concevoir l'interprétation suivante : les photons émis n'ont pas tous la même énergie : les un ont une fréquence ν_1, les autres une fréquence ν_2, etc., et le nombre de photons de fréquence ν_i est proportionnel à $I(\nu_i)$.

Mais si, comme nous l'avons fait, nous identifions chaque photon individuel avec un paquet d'ondes, l'interprétation la plus naturelle est la suivante : **le photon émis n'a pas, au moment de son émission, puis de sa propagation, une énergie bien définie. C'est seulement lorsqu'on cherche à déterminer son énergie (expérimentalement, sa longueur d'onde) que le photon est en quelque sorte obligé d'adopter une des énergies qui lui sont permises, et la probabilité qu'il a d'adopter l'énergie ν_i est proportionnelle à $I(\nu_i)$.**

Si l'on observe un grand nombre de photons, on retrouvera ainsi le spectre de fréquence $I(\nu)$.

Cette deuxième interprétation est l'interprétation probabiliste admise en mécanique quantique. Elle est due à Max Born et souleva autant de controverses que la théorie de la relativité. (Il faut, en particulier, signaler qu'Einstein n'en fut jamais tout à fait convaincu. Si vous doutez, vous êtes donc en bonne compagnie!)

La première interprétation est moins choquante : c'est notre insuffisance de connaissances qui ne nous permet pas de prévoir rigoureusement, et nous devons nous contenter d'une loi de probabilité. Dans l'interprétation probabiliste quantique, au contraire, le fait qu'on ne puisse pas, en général, prévoir

(1) Cette intensité est le carré de l'amplitude avec laquelle chaque fréquence intervient dans la décomposition de la vibration en somme de Fourier.

exactement le résultat d'une mesure, mais seulement calculer les probabilités respectives des différents résultats possibles, est dans la nature des choses.

la fonction d'onde

L'expérience des trous de Young a déjà été présentée en optique et interprétée grâce à la théorie ondulatoire de la lumière — c'est cette expérience qui donna un coup mortel à la théorie corpusculaire de Newton. Comment allons-nous interpréter cette expérience, maintenant que la lumière ne nous apparaît comme ni entièrement ondulatoire ni entièrement corpusculaire, mais comme formée de photons dont chacun peut se décrire par un paquet d'ondes? Nous allons voir en particulier que, présentés ainsi, les deux aspects, corpusculaire et ondulatoire, ne s'opposent pas, mais au contraire se complètent, ce qui permet une compréhension plus profonde du phénomène. On fait passer un faisceau parallèle de lumière par deux trous F_1 et F_2 très fins, et l'on observe sur un écran E la lumière transmise. En tous points A tels que la crête de l'onde issue de F_1 s'y superpose à la crête de l'onde issue de F_2, l'intensité lumineuse est maximale, alors qu'il y a extinction aux points B, où une crête se superpose à un creux. Dans le plan de l'écran E on observe des franges d'interférence.

La théorie ondulatoire de la lumière permet de calculer non seulement la position des franges d'interférence, mais la variation de l'intensité lumineuse sur l'écran d'observation — c'est le carré de ψ, somme des amplitudes des deux vibrations issues de F_1 et de F_2.

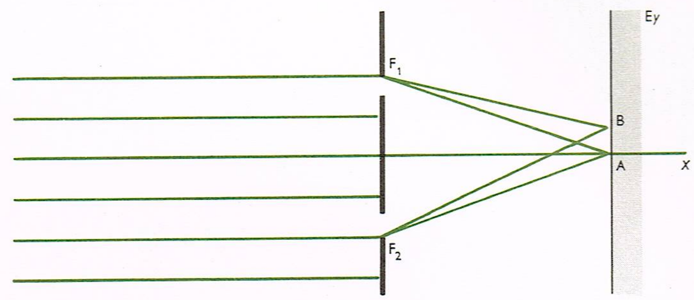
N'oublions pas que la lumière incidente n'est pas une onde plane, mais bien une grêle de photons. Quand on observe les franges au moyen d'un film photographique, par exemple, le noircissement d'un grain de l'émulsion est dû à l'action d'un photon individuel, qui, en cédant son énergie au grain de bromure, l'a rendu conducteur.

Si l'on pouvait réaliser un faisceau si peu intense que les photons arrivent un par un, à des intervalles de temps assez grands, sur un écran fluorescent, on observerait des scintillations successives bien localisées en différents points de l'écran. Si, au lieu d'un écran fluorescent, on utilise une plaque photographique, ces points d'impact s'accumulent, et si, après un temps assez grand pour que beaucoup de photons soient arrivés, on développe la plaque, on observe les célèbres franges d'interférence.

L'intensité de ces franges étant proportionnelle au carré de l'amplitude de l'onde lumineuse, l'interprétation probabiliste quantique est la suivante : **chacun des photons peut se matérialiser en n'importe quel point de l'écran, et la probabilité d'apparition de ce phénomène est proportionnelle au carré de l'amplitude ψ de l'onde qui lui est associée.**

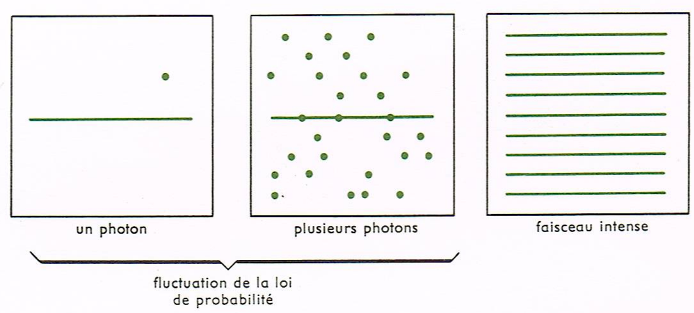
Nous aurions pu faire la même analyse pour les expériences d'interférences avec des faisceaux d'électrons, de neutrons, etc., et nous admettrons dorénavant l'interprétation probabiliste de la fonction d'onde ψ attachée à une particule ψ : le carré de ψ est proportionnel à la probabilité qu'a la particule de se trouver au point x, y, z, à l'instant t.

Schéma de l'expérience des trous de Young.

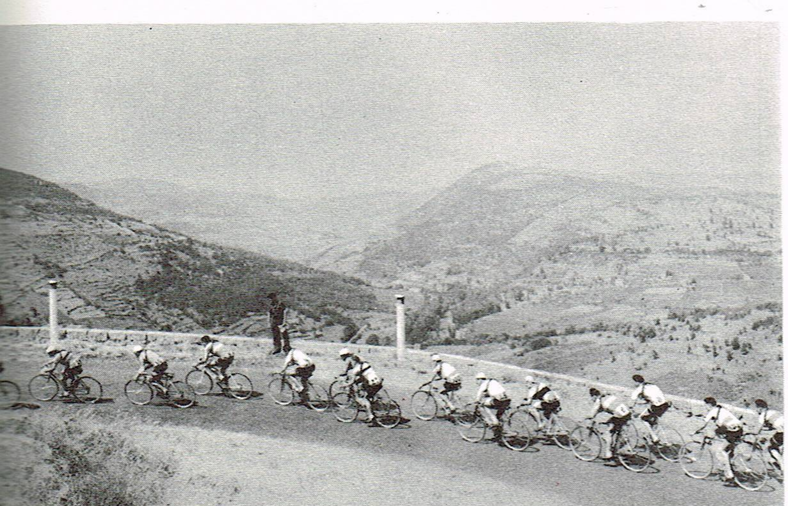


théorie ondulatoire de la lumière $\begin{cases} \text{si } F_2B - F_1B = \frac{\lambda}{2}(+n\lambda), \text{ extinction} \\ \text{si } F_2A - F_1A = 0\ (+n\lambda), \text{ maximum de lumière} \end{cases}$

Observation sur l'écran.



La loi de probabilité et la théorie ondulatoire donnent des résultats identiques.



Dans les épreuves cyclistes, chaque coureur a, du départ à l'arrivée, un numéro sur son dossard. Le fait qu'il soit parfois difficile de lire ce numéro ne provient pas d'une indétermination foncière, mais simplement d'une mauvaise méthode expérimentale (temps de pose trop grand).

Les photons d'une raie lumineuse sont différents : au départ, un photon individuel n'est pas affecté d'une longueur d'onde fixe, comme le coureur l'est d'un numéro. Lorsque, au moyen d'un appareil de mesure, on pose la question : « Photon, quelle est ta longueur d'onde? », le photon choisit parmi toutes les longueurs d'onde possibles, avec une probabilité de choisir une certaine longueur d'onde proportionnelle à l'intensité de celle-ci dans le spectre de la raie. C'est comme si les coureurs tiraient leurs numéros au sort au moment du sprint!

Plus généralement, la fonction ψ nous donnera, nous l'avons vu dans le cas de l'énergie du photon et nous verrons bientôt comment dans les autres cas, la probabilité qu'a toute grandeur physique attachée à la particule (son énergie, sa quantité de mouvement, son moment cinétique...) d'avoir une certaine valeur.

Dans l'expérience des trous de Young le phénomène n'est pas dû à l'interférence de l'onde représentant un photon avec l'onde représentant un autre photon; c'est l'onde représentant chaque photon individuel qui est responsable de cette loi de probabilité. On dit, de façon abrégée, que **chaque photon « interfère avec lui-même »**, et l'on est obligé d'admettre que le photon est passé par les deux trous à la fois. Cela paraît paradoxal, puisque le quantum d'énergie est indivisible. En fait, il n'y a pas de paradoxe, car, tant que nous n'essayons pas de savoir par quel trou le photon passe, le photon n'a pas à céder d'énergie, et si nous plaçons un scintillateur en F_1 pour déceler le passage du photon, nous détruisons la figure d'interférence. **Il n'y a pas opposition entre la description ondulatoire et la description corpusculaire, mais, comme l'a dit Bohr, ces deux descriptions sont complémentaires.**

En résumé, la fonction d'onde ψ contient toutes les informations que nous pouvons avoir sur un système. Le problème de la mécanique quantique se ramène donc maintenant à la détermination de ψ.

une limitation fondamentale de la connaissance : les relations d'incertitude

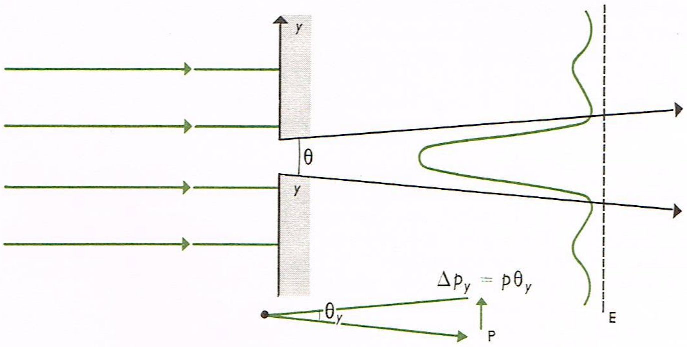
Auparavant, il nous faut insister sur une conséquence générale extrêmement importante de l'interprétation probabiliste de la fonction d'ondes. De façon quasi automatique, l'hypothèse ondulatoire entraîne une limitation fondamentale de la précision avec laquelle nous pouvons connaître simultanément la position et la vitesse d'une particule. Cette limitation est traduite par les célèbres relations d'incertitude d'Heisenberg.

Reprenons le cas du faisceau de photons. Nous avons vu que, pour donner une certaine localisation à la particule, il fallait en former la fonction d'ondes par une superposition d'ondes planes.

Dans notre interprétation, l'énergie du photon peut avoir une énergie quelconque $E = \hbar\omega$ autour de l'énergie centrale $E = \hbar\omega_0$ et dans une bande de largeur $\Delta E = \hbar \Delta\omega$; d'où la relation $\tau\Delta E \sim \hbar$ (puisque $\tau\Delta\omega \sim 1$).

De même, la longueur du train d'ondes — c'est-à-dire l'indétermination Δx sur la position x du photon à un instant donné — étant $c\tau$ (c, vitesse de propagation), et sa quantité de mouvement p étant connue avec une indétermination $\Delta p = \dfrac{\hbar\Delta\omega}{c}$, on a, entre ces deux indéterminations, la relation $\Delta x\, \Delta p \sim \hbar$, que nous avons écrite pour une dimension, et qui est, bien entendu, valable pour les deux autres. L'ensemble de ces trois relations et de la précédente est connu sous le nom de **relations d'incertitude d'Heisenberg** (1925) :

$$\tau\Delta E \sim \hbar \qquad \Delta x\, \Delta p_x \sim \hbar \qquad \Delta y\, \Delta p_y \sim \hbar \qquad \Delta z\, \Delta p_z \sim \hbar.$$



Passage de particules à travers une fente de largeur $2\Delta y$.

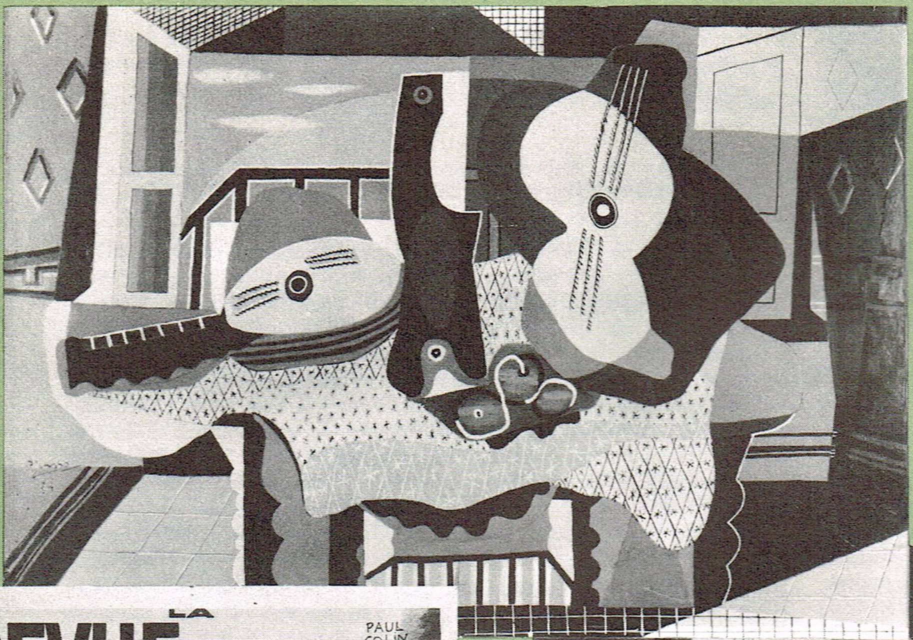
Nous venons d'en prouver la validité pour un cas particulier (mouvement en ligne droite d'un photon); on peut, à partir de l'équation de Schrödinger, les démontrer de façon très générale (1). Nous verrons même plus tard que les relations d'incertitude ne sont pas limitées aux grandeurs : positions (x) et quantité de mouvement (p).

La relation d'incertitude entre x et p signifie qu'il existe une limite fondamentale à la précision avec laquelle on peut déterminer simultanément la position et la quantité de mouvement d'une particule.

Par exemple, si l'on connaît exactement la quantité de mouvement ($\Delta p_x \to 0$), on ne peut absolument rien dire de la position ($\Delta x \to \infty$); c'est bien ce qui se passe lorsqu'on représente un faisceau monoénergétique par une onde plane monochromatique.

Cherchons de même à préciser la position suivant l'axe y des particules dans un faisceau qui se propage suivant l'axe Ox ($p_y = 0$). Pour cela, nous interposons un diaphragme percé d'une ouverture de largeur $2\Delta y$. Nous connaissons maintenant, à Δy près, l'ordonnée y des particules transmises. Mais le diaphragme produit un phénomène de diffraction; si l'on intercepte le faisceau par un écran, on observe non pas une tache de largeur Δy, mais une tache de diffraction beaucoup plus large; le faisceau, parallèle avant le diaphragme, a maintenant des rayons distribués à l'intérieur d'un cône de demi-ouverture $\theta_y = \dfrac{\lambda}{2\Delta y}$, où λ est la longueur d'onde de Louis de

(1) Ainsi présentées, les relations d'incertitude apparaissent comme une conséquence de la nature ondulatoire des particules. Dans la démarche d'Heisenberg, elles résultent immédiatement des principes de calcul qu'il pose. Schrödinger a montré que les deux méthodes (la sienne et celle d'Heisenberg) sont équivalentes; il n'y a donc pas lieu de rechercher lequel des deux phénomènes — nature ondulatoire ou principe d'incertitude — est le plus fondamental : ce sont plutôt deux façons de comprendre les choses, équivalentes quant au résultat. La notion de fonction est peut-être plus commode, parce que plus parlante, alors que, d'après Heisenberg, tout essai de représentation est en soi, sinon faux, tout au moins incontrôlable : la seule connaissance que nous puissions avoir est fournie par la mesure.



Pablo Picasso *Mandoline et guitare* **(1924).**



Un chien andalou **film de Luis Buñuel (1929).**



Affiche de Paul Colin (1925).



Le « Groupe des six », en 1922, au Bœuf sur le toit. *De gauche à droite,* **Darius Milhaud, Jean Cocteau (au piano), Arthur Honegger, Germaine Tailleferre, Francis Poulenc, Louis Durey. Georges Auric, absent, silhouetté par Cocteau.**

matrice de Pauli

$$\sigma_x = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \quad \sigma_y = \begin{vmatrix} 0 & -i \\ i & 0 \end{vmatrix} \quad \sigma_z = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$$

$$\sigma_z \left| +\frac{1}{2} \right> = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix} \begin{vmatrix} 1 \\ 0 \end{vmatrix} = \begin{vmatrix} 1 \\ 0 \end{vmatrix} = \left| +\frac{1}{2} \right>$$

$$\gamma_\mu \frac{\delta \psi}{\delta x_\mu} + m\psi = 0$$

équation de Dirac

$$\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = 2\delta_{\mu\nu}$$

Ci-dessous, ***le Cuirassé « Potemkine »*** **film de S. M. Eisenstein (1925).**



Les années 1920-1930 ont été marquées par une action révolutionnaire contre le classicisme dans le domaine de la pensée artistique et scientifique. Des concepts nouveaux se sont introduits; certains sont déjà morts, mais la mécanique quantique subsiste.

Phot. Giraudon, Cinémathèque française, Larousse, Lipnitzki.

Broglie ($\lambda = h/p$). A la sortie du diaphragme, les particules ont donc acquis une certaine quantité de mouvement p_y, perpendiculairement à leur direction initiale; de plus, cette quantité de mouvement, dite « quantité de mouvement transverse », n'est pas très bien définie : tout ce que l'on sait, c'est qu'elle est de l'ordre de $p\theta = \frac{\hbar}{2\Delta y}$.

On voit qu'en cherchant à préciser à y près la position transverse des particules, on a introduit une incertitude Δp_y sur la composante transverse de la quantité de mouvement, et qu'entre ces deux grandeurs on a nécessairement la relation

$$\Delta y \cdot \Delta p_y = \Delta y \cdot \frac{\hbar}{2\Delta y} = \frac{\hbar}{2},$$

qui est bien, à un facteur numérique près (1), l'une des relations d'Heisenberg.

Nous avons, pour cet exemple, raisonné par commodité sur un faisceau de particules. Mais, d'après l'interprétation statistique de la mécanique quantique, le même raisonnement est vrai pour une particule isolée : si nous obligeons un électron à passer par une fente de largeur Δy, la direction de propagation de l'électron après passage dans la fente est indéterminée à l'intérieur du cône défini ci-dessus. Quand on met un écran pour observer l'électron, l'électron est bien obligé de « choisir » une seule parmi toutes les valeurs θ possibles. Cependant, au moment où il est passé par la fente, et où l'on a donc déterminé sa position à Δy près, il est impossible de prévoir cette valeur de θ : tout ce que l'on peut dire, c'est qu'elle sera comprise entre $-\frac{\lambda}{2\Delta y}$ et $+\frac{\lambda}{2\Delta y}$.

Une « expérience de pensée » (Gedankenexperiment) due à Heisenberg illustre bien le principe d'incertitude, et montre que, même dans un cas très simple, on ne peut pas échapper aux conséquences de ce principe. Essayons de déterminer aussi précisément que possible la position d'une particule; pour cela regardons-la avec un microscope dont le pouvoir de séparation est donné par la relation bien connue :

$$\Delta y = 1{,}22\,\frac{\lambda}{\sin u}.$$

Mais, pour « voir » la particule, il faut qu'elle ait diffusé une partie de la lumière produite par le condenseur, donc qu'elle ait été frappée par au moins un photon, lequel s'est ensuite dirigé à l'intérieur du cône d'ouverture u. Au cours de ce choc la quantité de mouvement p_y de la particule a varié d'une manière opposée à celle du photon. Et l'on ne sait donc rien de la grandeur Δp_y de la quantité de mouvement échangée, sinon qu'elle est comprise entre 0 et $p \sin u$. L'incertitude Δp_y vérifie bien la relation $\Delta y \Delta p_y = \frac{1{,}22\lambda}{\sin u}\, p \sin u = 1{,}22\,\hbar \sim \hbar$.

Cette expérience et d'autres du même genre ont été beaucoup discutées, mais aucun perfectionnement n'a jamais permis de prendre en défaut la relation d'incertitude. Ainsi, on pourrait essayer de diminuer y en déterminant avec précision le centre de la tache due à la diffraction, mais, pour cela, il faut utiliser une lumière intense; cela implique que beaucoup de photons vont être diffusés par la particule, et augmente l'incertitude sur la quantité de mouvement de recul.

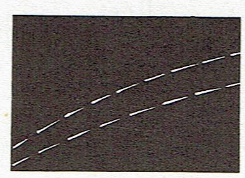
Les relations d'incertitude limitent donc de façon fondamentale notre connaissance de l'état d'un système; du fait que la détermination simultanée exacte de la position et de la vitesse d'une particule est impossible, le problème de la mécanique classique qui consiste à déterminer la trajectoire d'une particule à partir des données de position et de vitesse initiale n'a plus de sens.

(1) La forme exacte des relations d'incertitude est $\Delta x \Delta p_x \geqq \hbar/2$, l'égalité n'étant réalisée que dans des cas très particuliers.

inventaire des nouveaux concepts introduits

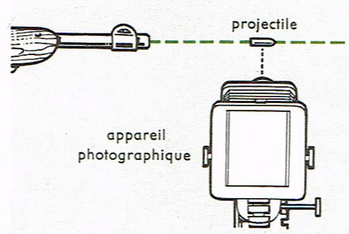
Nous sommes partis de la correspondance, somme toute assez simple, entre une particule en mouvement et une onde représentant la propagation d'une vibration. La validité de cette représentation est confirmée par les phénomènes d'interférence obtenus avec des électrons, des neutrons, des molécules. Toutefois, cette représentation est insuffisante pour localiser tant soit peu une particule en mouvement : il faut pour cela additionner un petit paquet de vibrations, de longueurs d'onde très voisines. Mais on s'aperçoit alors qu'on perd sur la connaissance de la vitesse ce que l'on gagne sur la connaissance de la position : ni l'une ni l'autre ne sont très bien définies, on ne peut donner pour chacune qu'une loi de probabilité. Ainsi la question « où est telle particule, à tel moment? » doit être remplacée par la question « quelle est la probabilité que la particule soit, à tel moment à tel endroit ? ». La réponse à cette question est donnée par le carré de la valeur qu'a la fonction d'onde à cet instant en cet endroit. De façon analogue, la question « quelle est la quantité de mouvement de la particule? » doit être remplacée par « quelle est la probabilité que la quantité de mouvement soit tant? », et sa réponse est donnée par l'intensité avec laquelle l'onde caractéristique de cette quantité de mouvement est représentée dans la fonction d'onde.

Le principe d'incertitude d'Heisenberg traduit simplement le fait que ces probabilités ne sont pas indépendantes (puisqu'elles se calculent à partir de la même fonction d'onde). Si l'on est à peu près sûr que x est compris dans un certain intervalle Δx, alors p doit être compris dans un certain intervalle Δp tel que $\Delta p\, \Delta x \sim h$. Telle est la limite que ne peuvent dépasser nos prédictions!

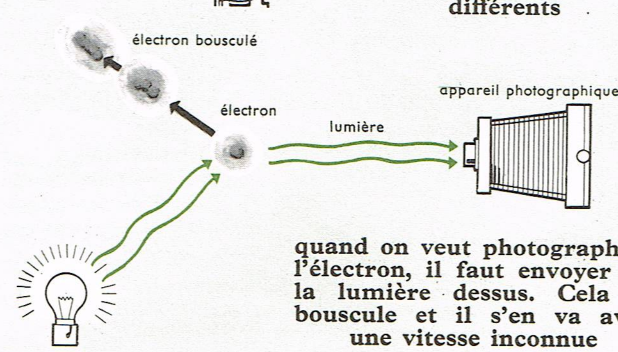


balles traçantes

Le principe d'incertitude, bien qu'universel, ne s'applique pratiquement qu'aux mouvements des particules atomiques.



on peut photographier le projectile à un moment donné. Ça ne change pas sa vitesse, que l'on peut calculer en mesurant l'intervalle de temps entre deux photos à deux endroits différents



quand on veut photographier l'électron, il faut envoyer de la lumière dessus. Cela le bouscule et il s'en va avec une vitesse inconnue

l'équation de Schrödinger

En mécanique classique, un système physique (atome, molécule, ensemble de particules) est décrit par les valeurs que prennent à chaque instant les positions et les vitesses des différentes particules. Du fait du principe d'incertitude, cela n'est pas possible en mécanique quantique : notre connaissance se réduit à ce que l'on peut tirer de la fonction d'onde. Il s'agit donc maintenant de pouvoir calculer cette fonction d'onde pour un système donné, puis, à partir de cette fonction d'onde, de tirer le maximum de renseignements sur les grandeurs physiques caractérisant le système — son énergie, son moment cinétique, etc. —, grandeurs que l'on appelle depuis Dirac les « observables ». Le premier problème — le calcul de la fonction d'onde — est résolu en principe par l'équation de Schrödinger propre au système. Le second — la détermination des observables — fera l'objet d'un chapitre ultérieur, chapitre où seront rassemblés les principes essentiels de la mécanique quantique.

Nous ne traiterons que le cas le plus simple, qui est aussi le plus général : la détermination de la fonction ψ décrivant un système « stationnaire », c'est-à-dire un système dont l'état ne dépend pas du temps. De tels systèmes existent dans la nature (par exemple, les états fondamentaux des atomes), ou encore sont de bonnes approximations pour les systèmes réels : la vibration d'une corde de violon, le mouvement d'un pendule quand on néglige les frottements, le mouvement des électrons autour d'un proton quand on néglige le rayonnement; de façon générale, tous les mouvements périodiques sont des exemples de systèmes « stationnaires ». Ainsi le mot « stationnaire » ne signifie pas que rien ne bouge, mais plutôt que le mouvement reste le même ou se reproduit identique à lui-même au cours du temps. Remarquons que tous ces systèmes ont ceci de commun que leur énergie est constante.

Un jet permanent de particules (représenté par une onde plane) constitue aussi un système stationnaire : rien ne distingue ce qui se passe maintenant de ce qui se passera plus tard. En termes quantiques : la probabilité de présence en un point donné du système, c'est-à-dire le carré de la fonction d'onde, est la même à tout moment; on dit qu'elle est indépendante du temps, et, de ce fait, il est en général inutile de faire figurer le facteur temps dans l'expression de la fonction d'onde d'un système stationnaire.

comment écrire le mouvement le plus simple, celui d'un électron en liberté

Pour le moment, la seule fonction ψ que nous connaissions est celle qui correspond à un faisceau parallèle monoénergétique de particules. Elle s'écrit en supprimant le facteur qui exprime la dépendance en fonction du temps : $\psi = \cos kx$.

Une opération mathématique simple — la double dérivation par rapport à x — montre que ψ est la solution de l'équation (où k est une constante) $\frac{d^2\psi}{dx^2} + k^2\psi = 0$.

C'est précisément l'équation à laquelle satisfait la propagation d'une vibration le long d'une corde, d'où son nom usuel d'« équation des cordes vibrantes ».

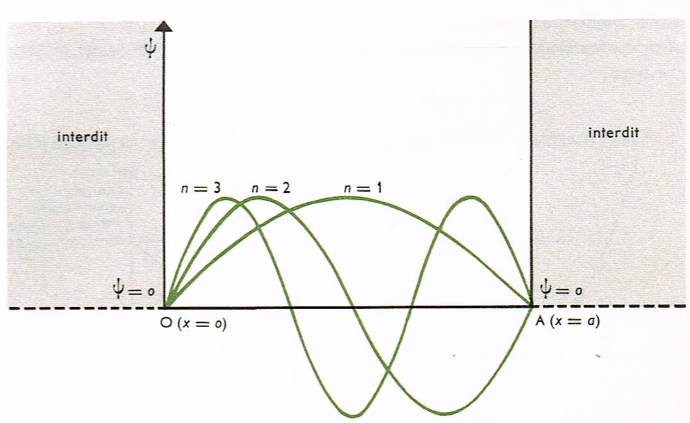
Réciproquement, l'équation précédente a pour solution la plus générale : $\psi = \cos(\pm kx - \varphi)$.

C'est-à-dire que toutes les ondes planes possibles se propagent suivant les deux directions de l'axe des x, avec le nombre d'ondes k. De plus, ces solutions existent quel que soit k, donc pour n'importe quelle quantité de mouvement. Or, à la question bien simple « quels sont les mouvements possibles d'une particule dans un espace à une dimension où ne règne aucune force? », nous aurions répondu « n'importe quel mouvement de translation rectiligne et uniforme ». C'est ce que donne, en termes d'onde, l'équation ci-dessus, qui est donc bien l'équation applicable à ce cas particulier.

Avant d'envisager le cas général d'une particule se déplaçant dans un champ de forces (par exemple dans un champ électrique ou magnétique, ou dans le voisinage d'un noyau), on peut dès maintenant appliquer cette équation à un problème bien simple, mais instructif. C'est le problème du mouvement libre d'une particule le long, non pas d'un axe infini, mais d'un segment de droite OA limité aux deux bouts. L'analogue classique est le mouvement d'une balle rebondissant entre deux murs verticaux. Plus généralement, on peut imaginer en O et A des forces répulsives assez grandes pour empêcher la particule de dépasser ces deux points (la représentation mathématique de telles forces sera donnée plus loin).

Donc, en dehors de OA la particule a une probabilité de présence ψ^2 rigoureusement nulle, et cela est vrai en particulier aux points O et A.

Par suite, parmi toutes les solutions sinusoïdales de l'équation, seules conviennent celles qui annulent ψ en O et A, ce qui impose une fonction ψ du type $\sin n\pi\frac{x}{a}$, où n est un entier; le segment OA comprend un nombre entier ou demi-entier de longueurs d'onde, la fonction $\sin\left(n\pi\frac{x}{a}\right)$ est bien nulle à l'extrémité gauche ($x = O$) et à l'extrémité droite ($x = a$). Ainsi, on a bien une onde d'amplitude nulle aux extrémités. On retrouve exactement les propriétés déjà connues d'une corde vibrante, avec un nœud (vibration nulle) à chaque extrémité; la corde



Les trois fonctions d'ondes les plus simples représentant les plus bas états d'énergie d'une particule entre deux murs (particule dans une boîte).

peut vibrer suivant le mode fondamental ($\lambda = 2a$, aucun nœud intermédiaire) et suivant tous ses harmoniques ($\lambda = \frac{2a}{n}$, $n - 1$ nœuds intermédiaires), elle ne peut pas vibrer n'importe comment.

deux notions fondamentales : l'« état » et la « fonction propre »

Ici, à chaque valeur de n correspond ce que l'on appelle un « état » de la particule, et la fonction $\psi_n = \sin n\pi\frac{x}{a}$ est la « fonction propre » qui représente cet état. On peut dire qu'une fonction propre décrit un mode particulier de vibration, c'est-à-dire un état. Dans chacun de ces états, la particule a une énergie bien déterminée $E_n = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} = n^2 \frac{\hbar^2}{8ma^2}$, **et aucune autre valeur de l'énergie n'est possible.** Voilà un résultat très différent de celui que donnerait la mécanique classique des particules : une balle peut rebondir entre deux murs avec n'importe quelle vitesse, donc n'importe quelle énergie, alors qu'ici existe une suite discrète d'énergies : E_1, E_2, ..., E_n...

C'est un résultat bienvenu, puisqu'une des tâches de la mécanique quantique est précisément d'expliquer pourquoi des systèmes tels que les atomes (appelés « systèmes liés ») ne peuvent avoir que certaines valeurs de l'énergie.

Quantitativement, la formule précédente donne une énergie cinétique minimale ($n = 1$) d'environ 35 électron-volts pour un électron évoluant à l'intérieur d'un segment de 1 Å (10^{-8} centimètre). C'est bien l'ordre de grandeur des énergies cinétiques des électrons dans les atomes légers (diamètre d'environ 10^{-8} cm).

l'équation de Schrödinger : un électron qui n'est plus en liberté

Jusqu'ici, seul le cas d'une particule libre a été abordé. Dans l'exemple précédent, à vrai dire, la particule n'est libre que dans l'espace restreint où elle n'est soumise à aucune force, alors qu'aux deux extrémités des forces intenses l'empêchent d'aller plus loin. Pour pouvoir traiter des exemples plus compliqués, il faut introduire dans une équation d'onde du type précédent quelque chose qui rende compte des forces qui s'exercent sur la particule.

Ces forces sont, par exemple :
— la pesanteur, force constante (à la surface de la Terre!) proportionnelle à la masse de la particule et de direction bien connue;
— la force électrique, proportionnelle à la charge et au champ électrique, constante dans un condensateur plan, ou variant comme $1/r^2$ s'il s'agit de l'attraction par une autre particule chargée (électron attiré par le proton dans l'atome d'hydrogène);
— la force élastique, proportionnelle à la distance dont la particule s'écarte de sa position d'équilibre, par exemple la force de rappel d'un ressort à boudin.

l'énergie potentielle

Les physiciens ont coutume de représenter la force $\vec{F}$ au moyen du potentiel V dont elle dérive. Ainsi :

Force nulle	$F = \vec{0}$,	donc potentiel constant	$V = V_0$
Force constante (par exemple pesanteur)	$\vec{F} = -\vec{mg}$	donc potentiel	$V = mgz$
Force de rappel d'un élastique (ou « oscillateur harmonique »)	$\vec{F} = -K\vec{r}$,	donc potentiel	$V = +\frac{Kr^2}{2}$
Force de Coulomb entre un électron et un proton	$\vec{F} = -\frac{e^2}{r^2}\frac{\vec{r}}{r}$,	donc potentiel	$V = -\frac{e^2}{r}$

Cela permet une représentation graphique simple et parlante de l'espace où évolue la particule.

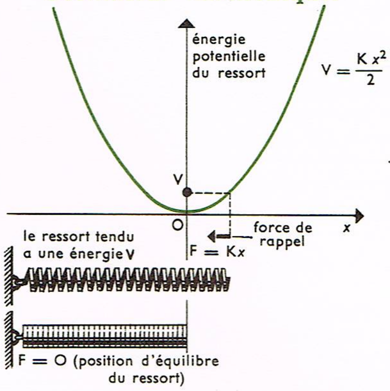
Outre cet avantage pratique, la notion de potentiel a une importance fondamentale en mécanique. En effet, à partir de l'équation de Newton $\vec{F} = m\vec{\gamma}$ (ou de sa forme relativiste), on démontre que la variation d'énergie cinétique dT d'une particule est égale et opposée à la variation de l'énergie potentielle dV, ce qui fait que dans un système isolé la somme $T + V$ est toujours constante. Ainsi une pierre tombant du haut de la tour Eiffel perd une énergie potentielle $g \times 300$ mètres et transforme cette énergie en une énergie cinétique ($1/2\ mv^2$) égale.

l'énergie de liaison et sa quantification

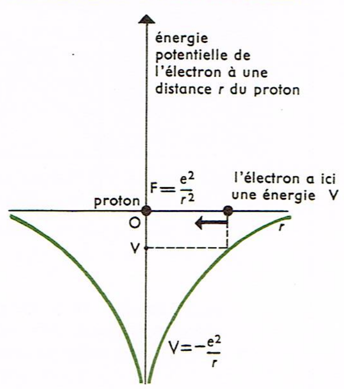
Examinons un des processus par lequel peut se former un atome d'hydrogène.

Un électron au repos ($T = 0$) à une distance très grande d'un proton ($V = 0$) constitue avec ce proton un système d'énergie initiale nulle. Bien que très éloignées l'une de l'autre, ces particules subissent une attraction électrostatique réciproque : l'électron se rapproche du proton (qui ne bouge pratiquement pas du fait de sa masse, 2 000 fois plus élevée que celle de l'électron).

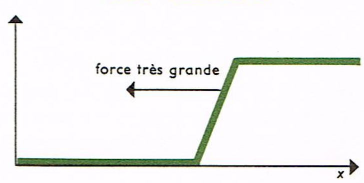
Oscillateur harmonique.



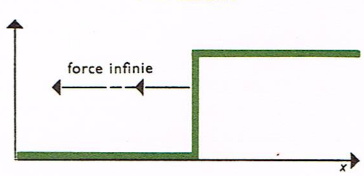
Attraction coulombienne



Une barrière.



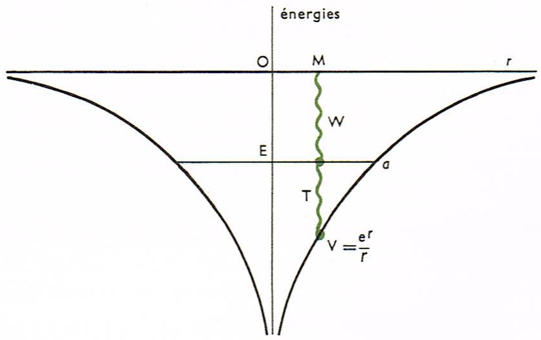
Un mur.



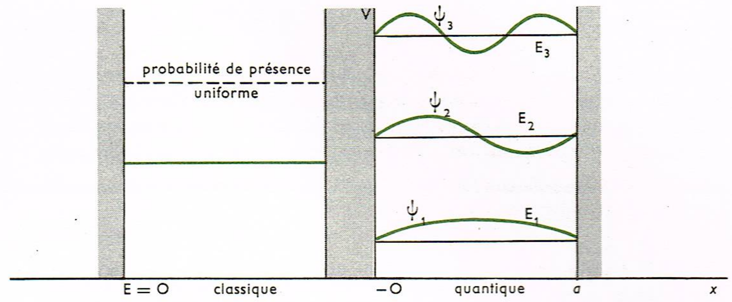
Exemples de potentiel

D'après la définition du potentiel (en vert), la force est toujours de sens opposé à la direction où le potentiel augmente plus rapidement (c'est-à-dire que la courbe représentant le potentiel est raide). A la limite, une force infinie en un point peut se représenter par un saut brutal du potentiel en ce point.

Énergie de liaison.



Particule entre deux murs. La fonction d'onde représentée en vert a exactement la forme d'une corde vibrante.



Au cours de ce rapprochement, l'énergie potentielle de l'électron va diminuer, son énergie cinétique augmenter comme pour la pierre tombant de plus en plus vite, jusqu'à ce que l'électron soit assez proche du proton pour être « capturé » et constituer avec le proton un atome d'hydrogène. L'électron a émis un rayonnement électromagnétique (puisqu'il a été accéléré), lequel a emporté une énergie W (positive); le système électron-proton, auquel nous pouvons par convention attribuer une énergie nulle lorsque les particules sont très éloignées l'une de l'autre, a donc maintenant une énergie totale E = — W (négative), que l'on appelle « énergie de liaison » (réciproquement, W = — E est l'énergie minimale qu'il faut fournir pour rompre le système). La loi bien connue et fondamentale de la conservation de l'énergie se traduit ainsi :

0	=	W	+	T	+	V,
énergie initiale du système		énergie dégagée par rayonnement de l'électron		énergie cinétique		énergie potentielle

ce qui donne pour l'énergie de liaison E = — W ou

E	**=**	**T**	**+**	**V**
énergie		**énergie cinétique**		**énergie potentielle**

Dans l'exemple précédent (problème de la particule dans une boîte), et contrairement à ce qui se passe en mécanique classique, toutes les valeurs de l'énergie de liaison ne sont pas possibles. Graphiquement, la restriction du mouvement au segment OA peut se représenter par deux barrières de potentiel de hauteur infinie, et la particule est « au fond » d'un « puits » infini de potentiel.

Dans cette représentation, les différents niveaux d'énergie constituent comme les barreaux d'une échelle — cela va se retrouver dans tous les systèmes liés, ainsi que d'autres propriétés telles que l'absence d'un état correspondant à la particule au repos et la corrélation entre le nombre de nœuds de la fonction d'onde et l'énergie du système.

Nous sommes maintenant assez avancés pour envisager d'écrire l'équation fondamentale de la mécanique quantique. Pour cela, commençons par reprendre l'équation d'onde de la particule libre telle qu'elle a été écrite plus haut : $\frac{d^2\psi}{dx^2} + k^2\psi = 0$.

Puisque $k = \frac{p}{\hbar}$, on peut écrire cette équation sous la forme $-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} = \frac{p^2}{2m}\psi$.

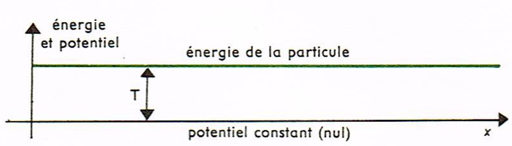
Ainsi nous faisons apparaître l'énergie cinétique de la particule $T = \frac{p^2}{2m}$. Supposons maintenant que la particule, au lieu de n'être soumise à aucune force, rencontre dans son déplacement le long de l'axe des x différentes zones où son énergie potentielle soit $V_0 = O$, V_1, V_2, ... Dans cette hypothèse, à l'intérieur de chacune des zones la particule peut être considérée comme libre, et l'équation précédente s'applique, à condition de prendre pour l'énergie cinétique $\left(T = \frac{p^2}{2m}\right)$ la valeur correcte, déduite de la conservation de l'énergie, T + V = E (énergie totale constante de la particule).

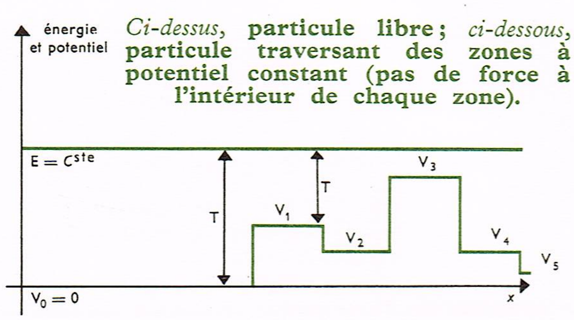
Nous voyons l'intérêt de récrire la relation sous la forme $-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} = (E - V)\,\psi$, ou encore $-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V\psi = E\psi$.

Il n'y a plus qu'un pas à faire, admettre que l'équation précédente est vraie quelle que soit la façon dont V varie en fonction de x, pour poser l'équation de Schrödinger à une dimension :

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V(x)\,\psi = E\,\psi.$$

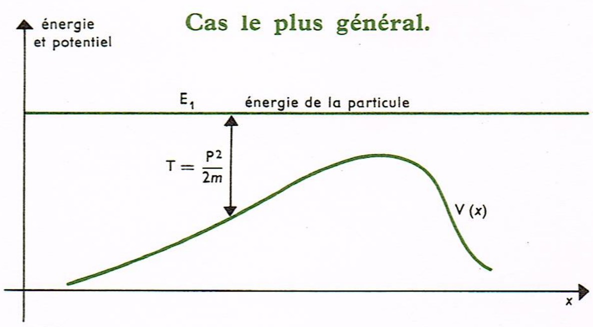
Telle est la célèbre **équation de Schrödinger;** indépendante du temps, elle ne s'applique qu'aux systèmes stationnaires (il existe une équation de Schrödinger dépendant du temps, qui permet de prévoir l'évolution des systèmes non stationnaires).





Ci-dessus, **particule libre;** *ci-dessous,* **particule traversant des zones à potentiel constant (pas de force à l'intérieur de chaque zone).**

Cas le plus général.



Trois diagrammes de potentiel représentant les forces qui s'exercent sur une particule au cours de son mouvement le long d'un axe.

Avant de passer à quelques exemples, notons bien que les lignes qui précèdent peuvent faire comprendre comment on a été amené à adopter, mais non certes à démontrer une formule tout aussi indémontrable directement que celle de Newton ($F = m\gamma$).

La validité de l'équation de Schrödinger est prouvée par le bon accord entre les prédictions que l'on peut en déduire pour les valeurs des grandeurs physiques caractérisant un système et les mesures expérimentales de ces grandeurs physiques. Certaines de ces prédictions sont étonnantes — et en contradiction avec ce que donnerait la mécanique classique —; l'expérience les a toujours vérifiées.

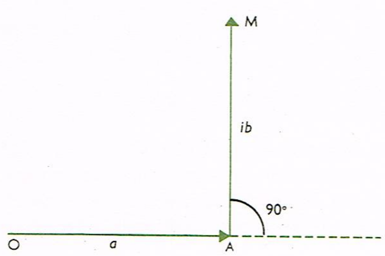
Auparavant, il nous faut revenir une dernière fois au plus simple des mouvements, celui de la particule libre, car cela va nous amener à une conséquence surprenante : l'introduction des nombres complexes (ou « imaginaires ») en physique.

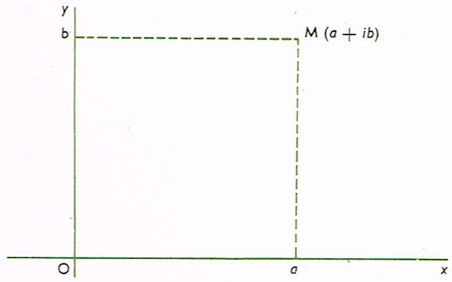
LES NOMBRES COMPLEXES ($a + ib$) ET L'EXPONENTIELLE $e^{i\varphi}$

Additionner deux nombres réels a (— 199, 5, — 1, 0, π, etc.) et b (— 47, 3, 0, 137, etc.), revient à mettre bout à bout deux segments, l'un OA de mesure a, l'autre AB de mesure b, et on obtient un segment OB de mesure $a + b$.

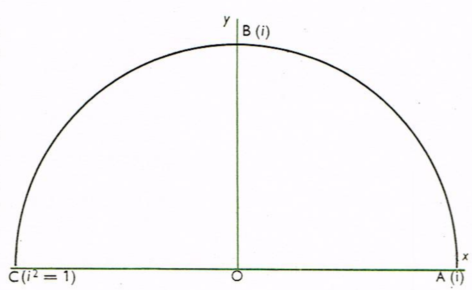


Mais si l'on fait tourner le segment AB de 90°, on obtient un point M différent de B. On peut convenir que ce point M représente un nombre z; par définition, a en est la partie réelle, et b la partie imaginaire, et l'on écrit : $z = a + ib$, où le symbole i représente la rotation de 90° qu'a subie le segment AB. On dit que z est un nombre complexe; il peut se représenter dans un plan xy (Ox est dit axe réel et Oy axe imaginaire).

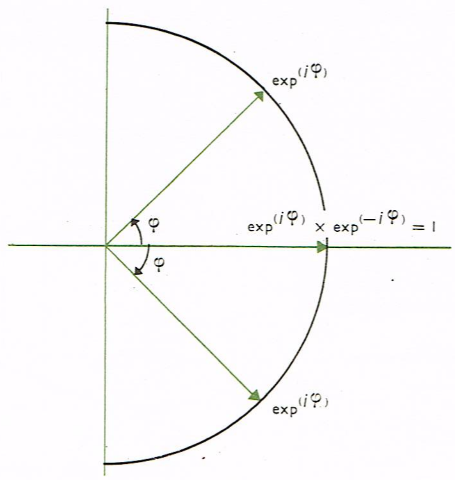




La multiplication d'un nombre par i consiste donc à faire tourner le segment représentant ce nombre de 90°. En particulier, si l'on considère le nombre réel unité (A), la multiplication par i donne le nombre i (B), une deuxième multiplication par i donne le nombre réel — 1(g). Ce qui conduit à écrire : $i^2 = -1$.



Le nombre a est réel. Le nombre ib est imaginaire pur. Le nombre $a + ib$ est imaginaire, le nombre $a - ib$ est le nombre imaginaire conjugué.



Une représentation particulièrement commode d'un nombre imaginaire z s'obtient en représentant le point M en coordonnées polaires, c'est-à-dire par la grandeur ρ de son rayon vecteur OM et l'angle φ de ce rayon avec l'axe réel; ce qui conduit aux formules

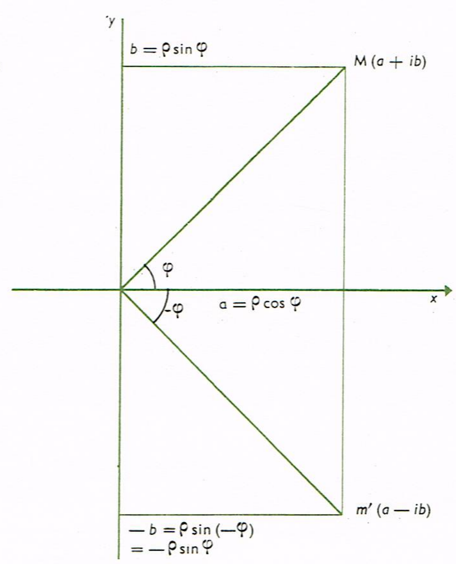
$$a = \rho \cos \varphi$$
$$b = \rho \sin \varphi$$
$$z = \rho (\cos \varphi + i \sin \varphi).$$

On peut encore poser :

$$\cos \varphi + i \sin \varphi = e^{i\varphi}$$

(en particulier $e^{\frac{i\pi}{2}} = i$),

en remarquant que la fonction ainsi définie (exponentielle de $i\varphi$) a les propriétés mêmes de l'exponentielle réelle.



On peut donc écrire le nombre complexe z et son imaginaire conjugué z* sous les deux formes équivalentes :

$$z = a + ib = \rho e^{i\varphi}$$
$$z^{\star} = a - ib = \rho e^{i\varphi}$$

Le produit

$$zz^{\star} = a^2 + b^2 = \rho^2$$

donne le carré de la grandeur du vecteur OM, c'est-à-dire le carré de la valeur absolue du nombre z.

nécessaire introduction de fonctions imaginaires

Nous avons décrit la propagation d'une onde plane par la fonction $\psi\ (x,\ t) = \cos\ (kx - \omega t)$.

Mais, si l'on y réfléchit un peu, cette forme n'est pas entièrement satisfaisante.

Si l'on observe, en effet, ce qui se passe en un point donné ($x = 0$, par exemple), on constate que la probabilité de présence $\psi^2 = \cos^2 \omega t$ fluctue entre 0 et 1; alors qu'un écoulement uniforme de particules monocinétiques, représenté précisément par l'onde plane, doit correspondre à une probabilité de présence constante en un point donné.

Il nous faut donc trouver une autre solution, correspondant à une probabilité constante. Problème d'apparence insoluble, mais fort simple si l'on fait appel aux mathématiciens, qui ont précisément inventé, il y a presque deux siècles de cela, et dans un but tout autre, l'outil même qui nous manquait : la fonction $e^{i\,(kx - \omega t)}$, qui jouit de la double propriété de satisfaire à l'équation $\frac{d^2\psi}{dx^2} + k^2\psi = 0$. et de correspondre à une probabilité de présence constante, et non fluctuante.

L'équation de Schrödinger peut donc avoir pour solutions des fonctions ψ complexes. Belle revanche accordée aux nombres qualifiés initialement d'imaginaires, parce qu'une au moins de leurs propriétés choquait le bon sens le plus commun ($i^2 = -1$)!

Bien entendu, s'il n'y a pas de difficulté mathématique à adopter un tel formalisme, il peut y en avoir au point de vue physique : que peut signifier une fonction d'ondes complexe, alors que toutes les grandeurs mesurées sont par définition réelles? La réponse est qu'il n'y a pas d'incompatibilité à décrire un système réel par une fonction ψ complexe, car la fonction ψ n'est pas directement observable. En revanche, il faudra que soient réelles les valeurs des grandeurs physiques mesurables, calculées à partir de ψ : les « observables ».

Un premier exemple en est donné par la probabilité de présence; quand ψ est réel, la probabilité de présence de la particule — nous l'avons déjà vu — est ψ^2, carré de l'amplitude ψ, ou encore carré de la valeur absolue ψ de cette amplitude (la probabilité de présence est la même pour les amplitudes $+\psi$ et $-\psi$).

Pour étendre cette définition à une fonction ψ complexe, il suffit de prendre le carré non de la fonction ψ, mais de sa valeur absolue $|\psi|$. On obtient bien ainsi un nombre réel $|\psi|^2$.

Cela équivaut à multiplier la fonction ψ, non par elle-même, mais par ce qui s'appelle la fonction « complexe conjuguée » ψ^*.

trois applications simples de l'équation de Schrödinger

Pour bien comprendre ces notions nouvelles, il est bon de revenir à l'exemple simple de la particule dans une boîte avant d'examiner deux problèmes un peu plus compliqués, pour lesquels l'utilisation de l'équation de Schrödinger est nécessaire et qui ont une grande importance théorique et pratique.

la particule enfermée entre deux murs

La solution de l'équation de Schrödinger avec les conditions aux limites $\psi = 0$ en O et A a été vue ci-dessus : $\psi = \sin \frac{n\pi x}{a}$, ce qui peut s'écrire $\psi = \frac{e^{\frac{inx\pi}{a}} - e^{\frac{-in\pi x}{a}}}{2i}$ c'est-à-dire que l'état « particule entre deux murs », appelé souvent « particule dans une boîte », est la superposition de deux ondes planes de quantités de mouvement égales et opposées. On retrouve l'analogie classique d'une balle rebondissant entre deux murs; la différence, déjà soulignée, est que seules certaines valeurs de l'énergie sont possibles.

Ce problème, apparemment très idéalisé, a d'importantes applications pratiques : grâce à la quantification des énergies, on peut calculer le nombre d'états possibles d'une particule évoluant dans un volume déterminé — c'est une des bases des statistiques quantiques.

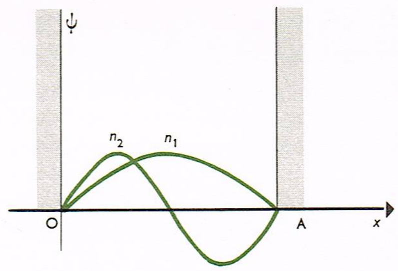
le franchissement des barrières de potentiel

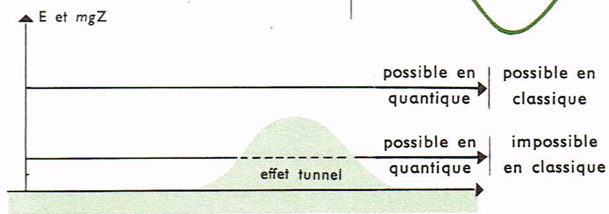
Le problème de la propagation de l'onde plane peut se compliquer lorsqu'on dresse un obstacle devant cette onde.

En mécanique classique, le problème est très simple. On peut considérer une skieuse essayant de franchir sur sa lancée une bosse de hauteur Z : si son énergie cinétique est plus petite que mgZ, elle est arrêtée en cours d'ascension; si elle est plus grande, elle franchit l'obstacle.

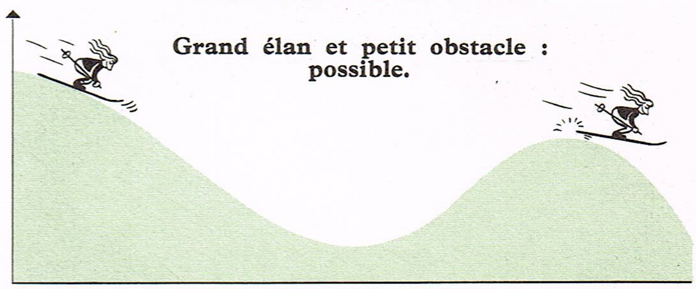
En mécanique quantique, l'analogue de la skieuse est une onde plane, et, pour simplifier la mathématique du problème, on

La particule entre deux murs.

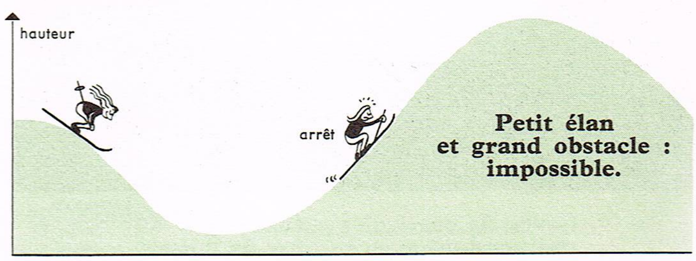




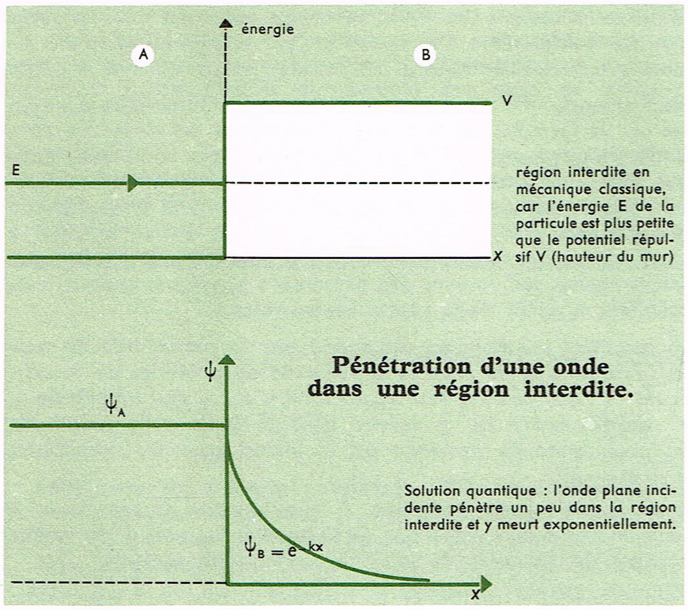
Franchissement d'une barrière.



Grand élan et petit obstacle : possible.



Petit élan et grand obstacle : impossible.

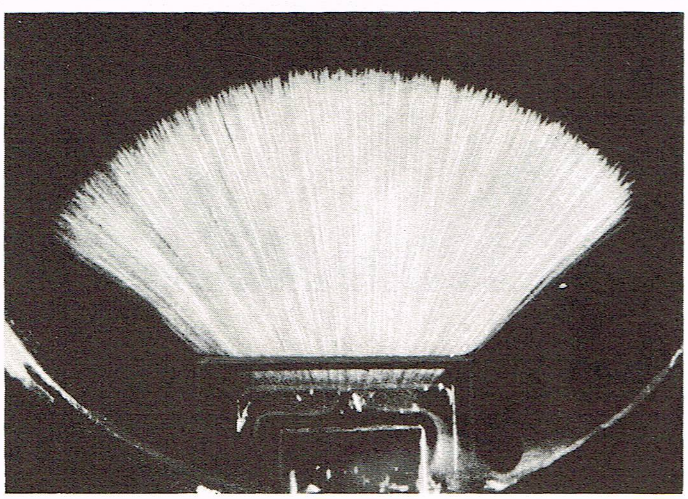


Pénétration d'une onde dans une région interdite.

peut d'abord remplacer la bosse par une falaise de potentiel de hauteur V.

Le cas le plus intéressant est celui où l'énergie E est plus petite que V. En mécanique classique, la particule rebondit alors sur le mur (la skieuse redescend la pente... comme elle peut!). Que va prédire la mécanique quantique? Ecrivons l'onde représentant la particule incidente (région A de la figure) $\psi_A = e^{i\frac{px}{\hbar}}$; p est la quantité de mouvement de la particule libre et s'exprime en fonction de son énergie cinétique $p = \sqrt{2mE}$.

Dans la région B, la solution de l'équation de Schrödinger s'écrit de la même façon, à condition de prendre pour l'énergie cinétique sa valeur T dans cette région. La conservation de l'énergie impose $E = T + V$, d'où une valeur négative pour l'énergie cinétique $T = E - V$. Cela est impossible en mécanique classique, et c'est pourquoi la skieuse ne franchit pas le mur. En mécanique quantique, nous écrirons que $p = \sqrt{2mT}$ a alors une valeur imaginaire pure :



Gerbe de particules α émises par du radium dans une chambre de Wilson.

$p = +\, i\sqrt{2m\,(V - E)} = +\, ihk$. Comme fonction d'onde possible nous avons alors $\psi_B = e^{i\,(+\,ik)\,x} = e^{-kx}$, c'est-à-dire une exponentielle décroissante.

L'existence d'une solution dans la région classiquement interdite entraîne, d'après la signification de la fonction d'onde ψ, une probabilité de présence ψ^2 non nulle à l'intérieur de cette région. **En termes plus simples, le faisceau de particules peut pénétrer dans la région interdite, pour y mourir rapidement** (ψ exponentielle décroissante). N'empêche que si petit que soit le domaine où ψ n'est pas nul, des particules peuvent exister dans une région où, classiquement, leur énergie serait insuffisante pour qu'elles y pénètrent. L'observation de ce phénomène va être une des preuves les plus frappantes de la validité de l'équation de Schrödinger.

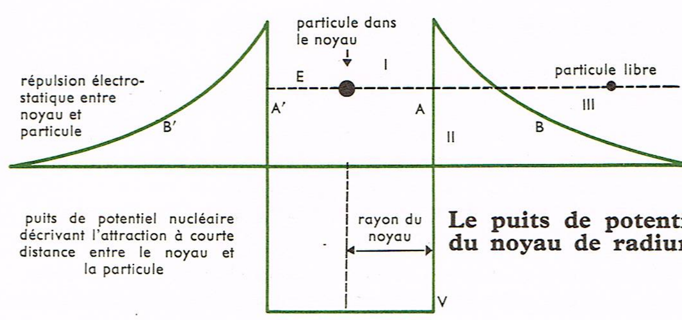
D'abord quelques ordres de grandeur sont nécessaires pour illustrer cette propriété de pénétration. Si la barrière culmine à 1 électron-volt au-dessus de l'énergie cinétique de la particule, la probabilité de trouver celle-ci à une distance de 2 angströms (10^{-8} centimètre) au-delà de la barrière est :
10 p. 100 (ou 0,1) pour un électron;
10^{-43} pour un proton.

L'électron a un pouvoir de pénétration non négligeable (de l'ordre des distances interatomiques existant dans un solide, ce qui expliquera la grande mobilité de certains électrons), alors que le proton est pratiquement arrêté par la barrière; cela tient uniquement au fait que le proton est 2 000 fois plus lourd. Point n'est besoin de calcul pour se rendre compte que cet effet purement quantique serait absolument inobservable pour des particules macroscopiques. Il n'y a donc aucune contradiction entre les faits expérimentaux courants (un cycliste descendant en roue libre le col d'Aubisque ne franchit jamais le Tourmalet sur sa lancée) et cette surprenante prédiction quantique.

Encore faut-il vérifier cette dernière; la nature en fournit différents exemples. Le plus célèbre est donné par la désintégration α de certains noyaux (radium, uranium, polonium) : radium → radon + particule α (noyau d'hélium). Ce phénomène a été la première des grandes manifestations nucléaires, observées dès la fin du siècle dernier (Becquerel, 1896). Il a fallu attendre trente ans pour que la mécanique quantique permette de le comprendre (Gamov, 1927).

comment les particules α peuvent s'échapper des noyaux

On s'imagine le noyau de radium comme constitué des noyaux de radon et d'hélium reliés entre eux par des forces attractives (nucléaires) très grandes; schématiquement, la particule α évolue à l'intérieur du puits de potentiel représentant cette attraction. Le rayon de ce puits est celui même du noyau de radium, soit 10^{-12} centimètre), et sa profondeur — s'exprimant



Le puits de potentiel du noyau de radium.

Cl. Philipp.

en millions d'électrons-volts — caractérise l'intensité des forces retenant la particule α dans le noyau. L'énergie de la particule α est positive, puisqu'elle peut s'échapper sans aide extérieure, et égale à son énergie à l'état libre (de 4 à 8 MeV suivant les émetteurs).

La courbe de potentiel se complète, à l'extérieur du noyau, par l'hyperbole $U = \frac{2Ze^2}{r}$ représentant la répulsion coulombienne entre le noyau final de radon (charge Ze) et le noyau d'hélium (charge 2e). C'est cette répulsion que le noyau d'hélium devrait vaincre si l'on voulait reconstituer le noyau de radium en bombardant du radon par des particules α.

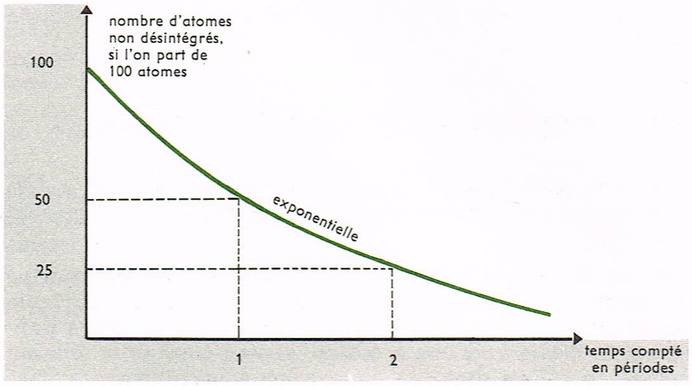
Dans l'état « radium », la particule α est à l'intérieur du puits de potentiel; comme « la particule dans une boîte », elle est représentée par une onde stationnaire, c'est-à-dire par la superposition de deux ondes, l'une se propageant de gauche à droite, l'autre de droite à gauche. Mais il y a quelques différences avec le problème de la particule dans la boîte.

L'une d'elles est que le mur de potentiel n'est pas infini, ce qui fait que la fonction d'onde pénètre en A, sous la forme exponentielle décroissante, dans la région « interdite » II.

Une autre différence est qu'à partir de B commence une région (III) où les conditions normales (énergie totale E supérieure à l'énergie potentielle V, c'est-à-dire énergie cinétique positive) sont de nouveau remplies. Si la particule atteint ce point B, elle va donc en quelque sorte sortir à l'air libre — et la désintégration du noyau de radium sera effectuée. Telle est, résumée brièvement, la théorie de Gamov. L'effet de pénétration au-delà de la barrière de potentiel est appelé, de façon imagée, **effet tunnel.**

des noyaux identiques ont une vie différente

La désintégration α fournit un autre exemple d'indétermination des phénomènes en mécanique quantique. L'indétermination dans l'espace nous est familière; ici, c'est l'indétermination dans le temps qui apparaît : étant donné un noyau de radium, on ne peut dire à quel moment il va se désintégrer, on peut simplement dire qu'il a 50 pour 100 de chances de le faire d'ici 1,6 an, 75 pour 100 de disparaître en 3,2 ans, etc.



Pourcentage d'atomes « survivant » au temps *t*. L'unité de temps est la période radio-active; après une période, il ne reste plus que 50 p. 100 des atomes; la période suivante, 50 p. 100 des atomes subsistants se sont désintégrés à leur tour, etc.

Plus généralement, un système excité a une certaine probabilité par unité de temps de se désintégrer — et l'on ne peut en prédire plus.

En toute rigueur, cette probabilité de transition se calcule par l'équation de Schrödinger dépendante du temps; c'est ainsi que la mécanique quantique rend compte des transitions entre états excités des atomes : intensité relative des différentes raies, largeur de ces raies — chose inaccessible à la théorie de Bohr.

Dans le cas de la radio-activité α, la théorie de Gamov permet de calculer cette probabilité — c'est-à-dire la vie moyenne de chaque espèce radio-active. Pour qu'il y ait désintégration, il faut qu'une particule α heurte la barrière de potentiel, puis qu'elle la traverse; ainsi, nombre de chocs × probabilité de transmission par « effet tunnel » = probabilité de désintégration.

L'accord avec l'expérience est excellent, qu'il s'agisse du radium, du thorium, de l'uranium ou du polonium. Ce résultat est d'autant plus remarquable que les vies moyennes de ces corps s'étagent entre des fractions de seconde et des milliards d'années.

l'oscillateur harmonique (le ressort à boudin quantique)

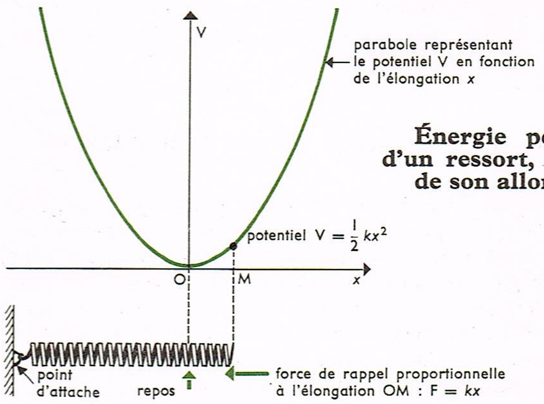
L'exemple de la particule maintenue à l'intérieur d'une « boîte » par deux barrières de potentiel infini est un cas limite. En général, une particule est liée à un centre d'attraction par l'intermédiaire d'un champ de forces plus ou moins compliqué, que l'on peut représenter par une courbe de potentiel. Ainsi le potentiel dit « de l'oscillateur harmonique », $V = \frac{1}{2} kx^2$, entraîne l'existence d'une force de rappel vers l'origine, proportionnelle à l'élongation (force élastique, comme dans le ressort à boudin).

Le problème est de trouver, pour une particule soumise à un tel potentiel, les fonctions d'onde possibles et les niveaux d'énergie correspondants.

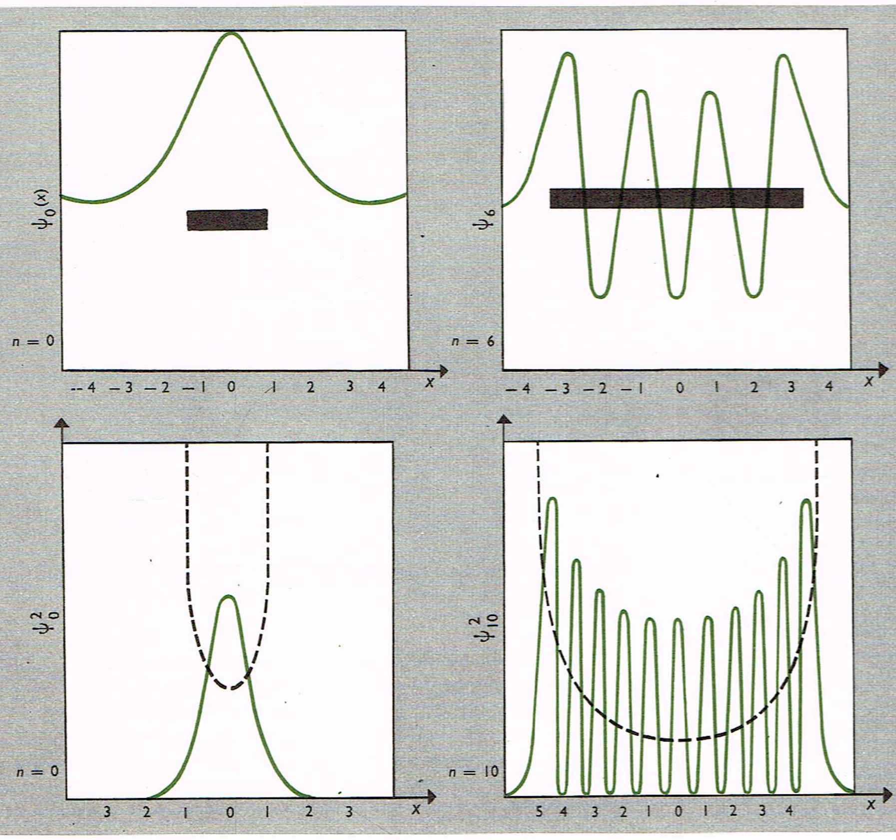
Ce problème n'est pas tellement différent de celui de la particule entre deux murs : si les conditions aux limites sont autres (il ne s'agit plus d'un arrêt brutal, mais d'une atténuation progressive du mouvement), le résultat reste cependant le même; l'équation de Schrödinger n'a de solution que pour certaines valeurs E_i de l'énergie (valeurs propres) formant une suite discrète (spectre), et à chaque valeur du nombre quantique i correspond une fonction d'onde (état propre).

On retrouve les propriétés des cordes vibrantes, pour lesquelles la fréquence du son émis ne peut prendre que certaines valeurs — cela étant toujours dû aux « conditions aux limites » (raccordement entre deux régions et nécessité pour les fonctions d'onde de s'annuler à grande distance).

De la même façon, **l'équation de Schrödinger introduit en quelque sorte automatiquement la quantification de**



Énergie potentielle d'un ressort, en fonction de son allongement.



Représentation quantique du mouvement de l'oscillation harmonique.

Les deux courbes du haut montrent 2 des fonctions d'onde possibles. Ces fonctions s'étendent hors des limites figurées par les barres en noir, dont les deux extrémités représentent les positions extrêmes du ressort classique.

Les deux courbes du bas représentent les carrés des fonctions d'onde et donc la probabilité de présence. La courbe en pointillé représente la même probabilité pour le mouvement classique du ressort.

l'énergie d'un système lié : elle n'a de solution que pour certaines valeurs de l'énergie, appelées « valeurs propres ». Les fonctions ψ correspondantes sont dites « fonctions propres ». On range d'habitude les valeurs E_i par ordre croissant E_1, E_2, ..., E_n, et n est alors dit « nombre quantique d'énergie ».

Revenons à l'oscillateur harmonique. C'est un cas particulier important, car il constitue souvent une excellente approximation du mouvement. Par exemple, on peut considérer qu'une particule liée à un système exécute autour de sa position d'équilibre un mouvement oscillatoire.

La résolution mathématique de l'équation de Schrödinger donne pour l'énergie E les valeurs propres suivantes :

$$E_n = (n + 1/2)\ h\nu,$$

où ν est la fréquence fondamentale d'oscillation que donnerait la mécanique classique.

Les fonctions propres ψ correspondantes sont représentées sur la figure jointe, où le trait horizontal épais représente le segment BC (amplitude du mouvement permis d'un oscillateur classique).

Niels Bohr et le principe de correspondance

La figure ci-dessus représente en traits pleins la probabilité de présence ψ^2, et en pointillés la probabilité de présence classique (inversement proportionnelle à la vitesse v de la particule, c'est-à-dire infinie aux deux points extrêmes où $v = 0$, et minimale au centre où v est maximal). On voit que pour $n = 0$ (état fondamental), la description quantique est très différente de la description classique, alors que pour $n = 10$ la courbe quantique est très proche de la courbe classique.

Ce graphique illustre une propriété générale connue sous le nom de « principe de correspondance »; **alors que, pour les petits nombres quantiques, la description quantique du système est très différente de la description classique, cette différence tend à s'atténuer lorsque le nombre quantique augmente,** c'est-à-dire lorsque l'énergie du système devient grande par rapport à la différence des énergies des états quantiques voisins.

Les propriétés physiques du système deviennent alors à peu près les mêmes que celles que prédirait la mécanique classique. Il faut bien, en effet, qu'à un certain niveau la mécanique classique devienne d'une bonne approximation, puisque, à notre échelle, celle-ci est excellente. Ce principe de correspondance a été un des guides de Bohr pour l'élaboration de sa théorie de l'atome.

une conséquence du principe d'incertitude

Une remarque s'impose : parmi les solutions $E_n = (n + 1/2)\ h\nu$ ne figure pas l'énergie zéro; l'énergie minimale $1/2\ h\nu$ n'est pas nulle.

Cela se comprend immédiatement à la lumière du principe d'incertitude : l'énergie zéro correspondrait à une particule rigoureusement immobile, dont on connaîtrait à la fois la position et la vitesse avec une précision infinie (toutes deux seraient nulles), ce qui est contraire au principe d'incertitude.

Historiquement, la quantification des énergies de l'oscillateur harmonique a été posée *a priori* par Planck, dès 1900, pour expliquer le spectre du rayonnement du corps noir. Il fallut attendre la théorie de Bohr pour rendre cette hypothèse compréhensible, et l'équation de Schrödinger pour l'expliquer pleinement.

les spectres des atomes enfin expliqués

Après ces exemples simples et quelque peu idéalisés, quoique réels, tournons-nous maintenant vers le phénomène concret qui a motivé l'introduction de la mécanique quantique : les spectres de raies des atomes.

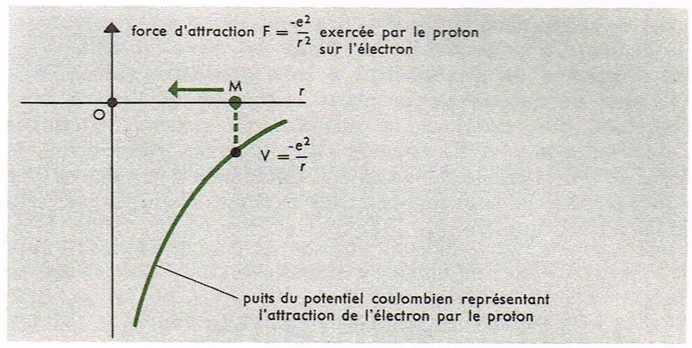
Nous n'étudierons guère ici que le plus simple des atomes : l'atome d'hydrogène. Cela suffira pour voir comment la mécanique quantique explique le mystère des niveaux d'énergie des atomes.

L'atome d'hydrogène est composé d'un proton et d'un électron, reliés entre eux par l'attraction électrostatique.

L'électron étant environ 2 000 fois plus léger que le proton, on peut considérer ce dernier comme fixe, et le mouvement de l'électron M se décrit comme le mouvement d'une particule autour d'un point fixe O, qui l'attire suivant la loi de Coulomb

$$F = \frac{e^2}{r^2}.$$

La différence essentielle avec les exemples précédents est que l'électron peut se déplacer dans tout l'espace et non seulement le long d'un axe. Il faut donc décrire le mouvement en fonction de trois coordonnées spatiales, et non d'une seule. Cela va introduire une difficulté mathématique; mais, heureusement, dans le cas de l'atome d'hydrogène la force ne dépend que de la distance, ainsi que le potentiel, qui est dit de ce fait « potentiel central ». Dans ce cas, le problème se simplifiera.



Le potentiel ne dépend pas de la direction dans l'espace du rayon OM, le long de n'importe quel axe passant par O; le puits du potentiel où évolue l'électron a la même représentation.

une première série de solutions

Il est évident que, lorsque les données du problème ne dépendent que de la distance (problème dit « à symétrie sphérique »), il doit exister des solutions qui ne dépendent, elles aussi, que de la distance, c'est-à-dire qui présentent la même symétrie.

Pour rechercher ces solutions, on posera donc *a priori* que la fonction ψ ne dépend que de r (distance de l'électron au proton), et, à partir de l'équation de Schrödinger à trois dimensions (1), on écrira l'équation reliant la solution $\psi(r)$ au potentiel $V(r)$. On retombera ainsi sur une équation du même type que l'équation de Schrödinger à une dimension.

La fonction ψ doit répondre à certaines conditions; en particulier, devenir très petite à grande distance du centre, ce qui traduit le fait que l'électron est attaché au proton. De la même façon que dans les problèmes de la particule dans une boîte ou de l'oscillation harmonique, l'existence de cette « condition aux limites » entraîne la quantification de l'énergie.

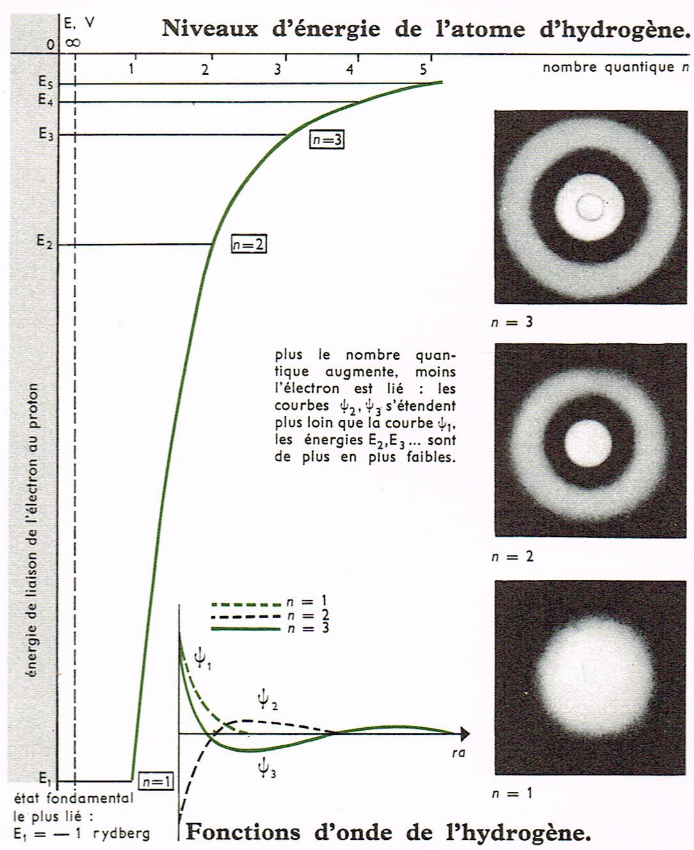
Un peu de mathématiques donne pour l'énergie les valeurs propres suivantes, qui reproduisent exactement les valeurs expérimentales trouvées plus de quarante ans auparavant par Balmer :

$$E_1 = -1 \text{ rydberg} = -13{\cdot}58 \text{ électron-volts};$$

$$E_2 = -\frac{1}{4} \text{ rydberg};$$

$$E_n = -\frac{1}{n^2} \text{ rydberg}.$$

Niveaux d'énergie de l'atome d'hydrogène.



Fonctions d'onde de l'hydrogène.

(1) Par simple généralisation de l'équation à une dimension

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + (E - V)\psi = 0,$$

on écrit l'équation à trois dimensions

$$-\frac{\hbar^2}{2m}\left[\frac{\delta^2\psi}{\delta x^2} + \frac{\delta^2\psi}{\delta y^2} + \frac{\delta^2\psi}{\delta z^2}\right] + (E - V)\psi = 0,$$

dont la solution est fonction en général des trois coordonnées x, y, z.

Le rydberg, rappelons-le, a pour valeur $E = \frac{me^4}{2h^2}$, valeur déjà trouvée par la théorie de Bohr, et en remarquable accord avec l'expérience.

Ainsi, la mécanique quantique, sans aucun artifice, rend compte très simplement des niveaux d'énergie de l'atome d'hydrogène.

mais comment peut-on essayer de se représenter l'atome?

Puisque le carré de la fonction d'onde représente la probabilité de présence, l'électron peut être n'importe où dans l'espace, avec — phénomène curieux — une probabilité maximale d'être au centre. **La belle orbite de Bohr est devenue une sorte de « sphère de probabilité » aux contours assez flous.** On peut néanmoins parler d'une dimension de l'atome : en calculant, par exemple, la distance du proton à laquelle l'électron a le plus de chances de se trouver. Ainsi défini, le « rayon » de l'atome dans son état fondamental se trouve être égal au rayon de la première orbite de Bohr, $a = 0{,}53 \cdot 10^{-8}$ centimètre (0,53 angström).

d'autres solutions, plus compliquées

Les états de l'atome décrits ci-dessus — états dont la fonction d'onde ne dépend que de la distance au centre, et dits, suivant la terminologie des spectroscopistes, **« états *s* »** — ne sont pas les seuls possibles.

Au lieu de se contenter de rechercher des fonctions de la distance seulement, on peut essayer de voir s'il y a aussi des solutions d'un type un peu plus compliqué, par exemple $xg(r)$. Non seulement il en est bien ainsi, mais on trouve toute une série de fonctions $g(r)$ possibles, correspondant chacune à une valeur propre de l'énergie :

$$E_2 = -1/4 \text{ rydberg};$$
$$E_n = -1/n^2 \text{ rydberg}.$$

Une curieuse particularité mathématique, propre au potentiel coulombien, fait que ces niveaux d'énergie reproduisent ceux de l'état *s*, sauf le premier.

Ainsi, la même énergie peut être obtenue avec deux fonctions d'onde différentes (et l'on peut même encore trouver deux autres fonctions d'onde, car rien ne distingue l'axe *x* des axes *y* et *z*). Pour l'énergie E_2, par exemple, il y a donc quatre fonctions d'onde possibles :

une fonction $f(r)$,
les fonctions $xg(r)$, $yg(r)$, $zg(r)$.

La première décrit l'état à symétrie sphérique « 2*s* », les trois autres constituent trois descriptions indépendantes d'un autre état, dit **« état 2*p* »** (toujours la notation des spectroscopistes).

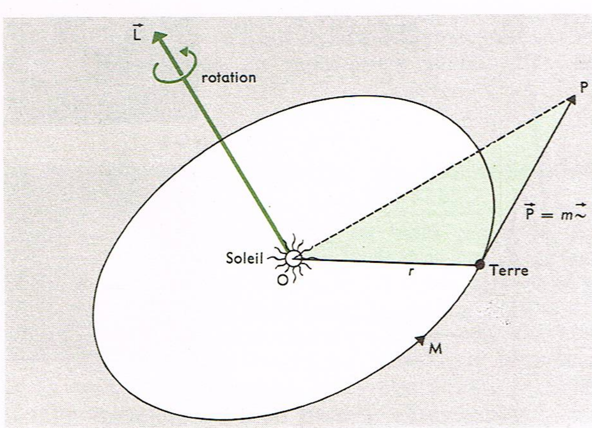
En quoi l'état 2*s* diffère-t-il de l'état 2*p*? L'application des règles de la mécanique quantique nous permettra bientôt d'énoncer de façon précise les caractéristiques de ces deux états. Pour le moment, la comparaison avec la théorie de Bohr peut nous permettre d'anticiper sur ce résultat.

retour à la théorie de Bohr

En effet, dans la théorie de Bohr de l'atome d'hydrogène, pour la même énergie, c'est-à-dire le même nombre quantique principal ($n = 2$, dans notre exemple), on peut trouver des orbites d'excentricités différentes (au nombre de *n*). Ces diverses formes d'orbite correspondent chacune à une valeur bien définie du moment cinétique.

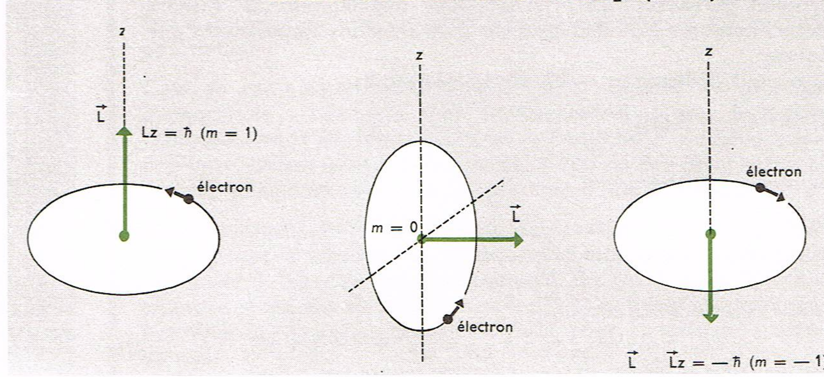
Rappelons qu'en mécanique classique le moment cinétique d'une particule évoluant dans un champ central est constant (dans le cas du mouvement des planètes autour du Soleil, cela se traduit par la célèbre « loi des aires » de Kepler [v. page 40]). Cette propriété se conserve dans la théorie de Bohr, où, de plus, le moment cinétique se trouve être quantifié, c'est-à-dire qu'il ne peut avoir que certaines valeurs ($0, \hbar, ..., [n-1]\hbar$). Si $n = 2$, par exemple, les moments cinétiques possibles 0 et $\hbar$ correspondent respectivement à un mouvement en ligne droite et à une trajectoire elliptique.

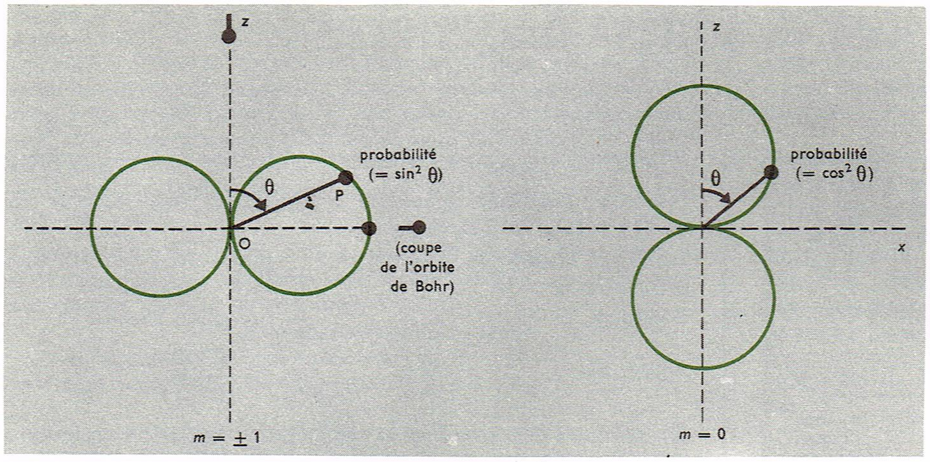
Pour définir complètement le mouvement, il faut, de plus, placer les orbites dans l'espace — cela peut se faire en trouvant la direction du moment cinétique, qui est perpendiculaire au plan de l'orbite. Ici intervient encore un phénomène de quantification : alors qu'en mécanique classique toutes les orientations sont possibles, si l'on applique les règles de Bohr pour trouver la direction du moment cinétique par rapport à un axe arbitraire — c'est-à-dire sa projection sur cet axe —, on trouve que cette projection elle-même est quantifiée : les valeurs possibles sont $m\hbar$, où *m* est un entier compris entre $-l$ et $+l$. Dans le cas de l'état *p* ($l = 1$), il y a donc trois projections possibles : $-\hbar$, 0 et $\hbar$.



Le moment cinétique est un vecteur $\vec{L}$ perpendiculaire au triangle formé par le rayon vecteur OM et la quantité de mouvement MP. Sa longueur est égale au double de l'aire de ce triangle. Dans le mouvement des planètes autour du soleil, l'aire du triangle est constante (loi de Kepler), le moment cinétique est perpendiculaire à l'orbite et est constant.

Atome de Bohr dans l'état *p* ($l = 1$).



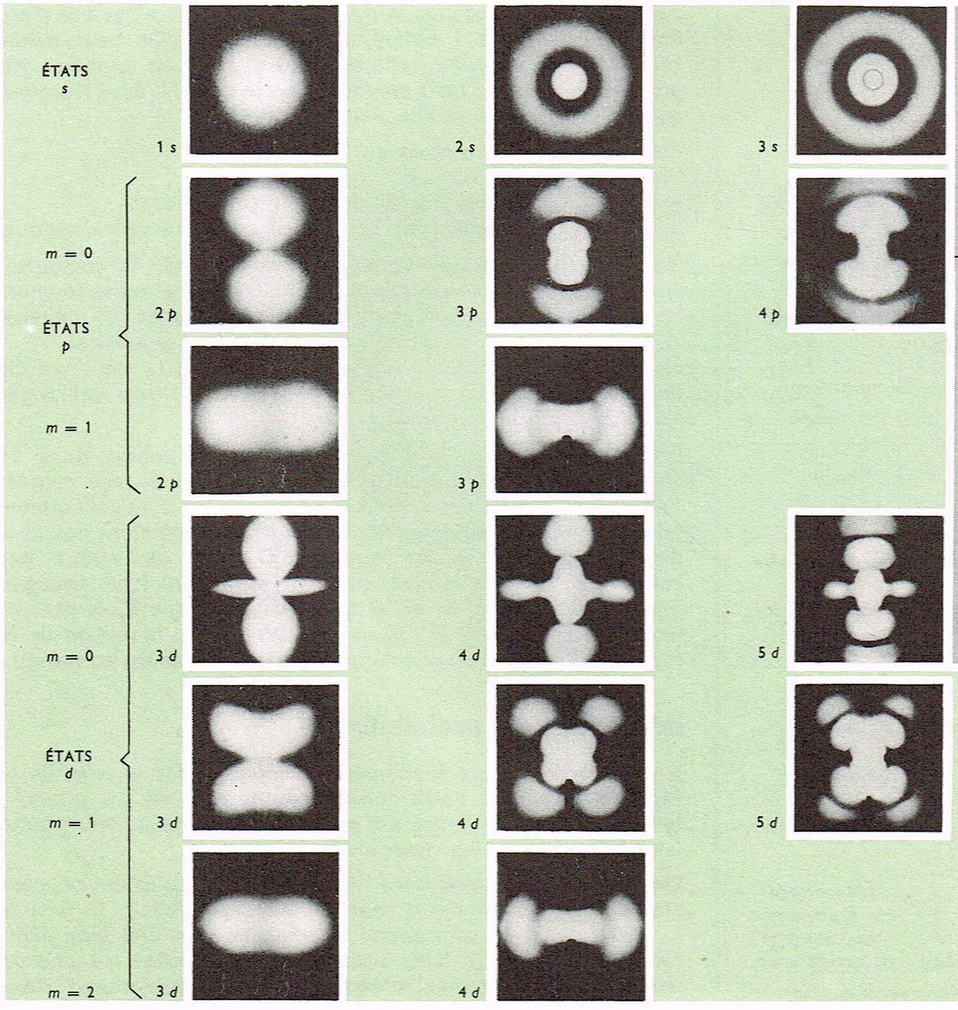


Probabilité de présence de l'électron dans les différentes directions.

La figure représente un plan quelconque passant par l'axe Oz, qui est un axe de révolution pour les états représentés. Dans ce plan, la probabilité pour trouver l'électron dans la direction d'angle zénithal θ (mais à n'importe quelle distance du proton) est proportionnelle au segment OP.
Pour $m = \pm 1$, le plan de l'orbite de Bohr est perpendiculaire à Oz.
Pour $m = 0$, le plan de l'orbite contient Oz.

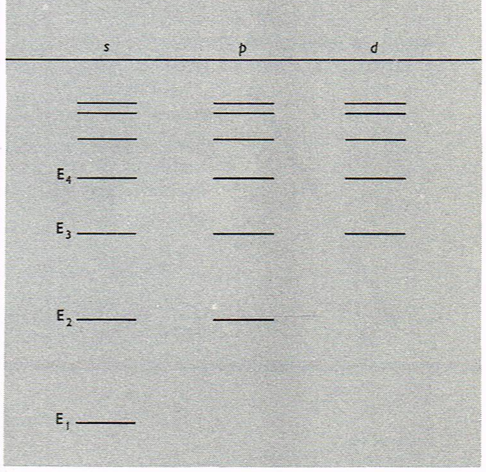
L'essentiel de ces résultats surprenants de la théorie de Bohr se retrouve en mécanique quantique : les états **s**, à une seule fonction d'onde, sont des états de moment cinétique nul (et par suite à une seule valeur, nulle, de la projection), les états *p*, à trois fonctions d'onde indépendantes, sont des états de moment cinétique $l = 1$ (donc à trois valeurs $m = -1, 0$ ou $+1$ de la projection) : les directions ne sont pas quelconques, mais sont elles aussi soumises aux lois de la quantification.

En revanche, la description ancienne, en termes d'orbites bien définies, doit être abandonnée, malgré sa commodité; ainsi, dans l'état $l = 1$, $m = 1$, l'électron décrirait, d'après Bohr, une ellipse dans le plan *xy*. En mécanique quantique, il peut se trouver sur n'importe quelle direction de l'espace, avec une certaine loi de probabilité, illustrée par un diagramme en « lobes » (toutefois, lorsque les nombres *l* et *m* sont grands, les lobes s'aplatissent et l'on se rapproche de la notion d'orbite — toujours le principe de correspondance).

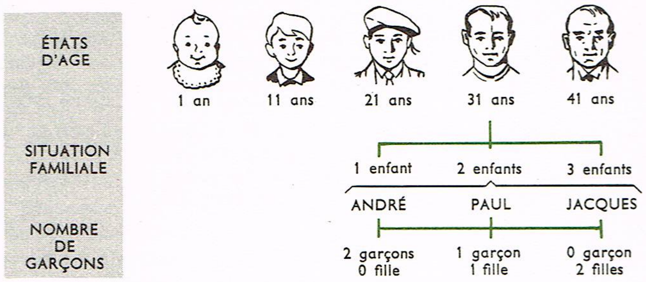


Nuages de probabilité représentant différents états de l'atome d'hydrogène.

Diagramme des niveaux d'énergie de l'atome d'hydrogène.

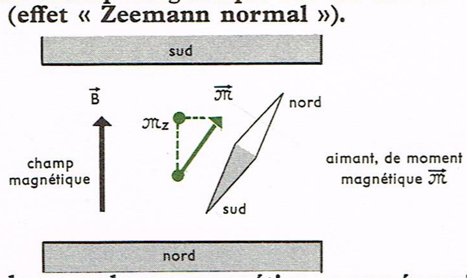


qu'appelle-t-on « états dégénérés » ?

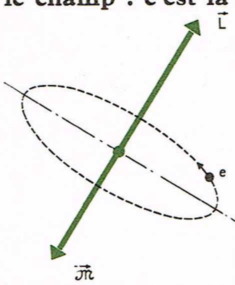


Jacques (31 ans, 2 filles), André (31 ans, 2 garçons), Paul (31 ans, 1 garçon et 1 fille) sont tous trois dans le même état d'âge, et de même pour l'état du nombre d'enfants, mais dans des états différents quant aux sexes des enfants.

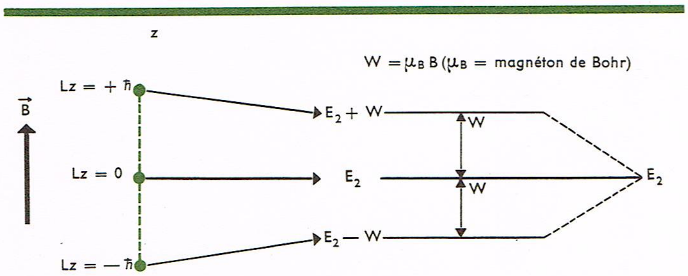
Effet d'un champ magnétique sur un atome (effet « Zeemann normal »).



Un aimant dans un champ magnétique a une énergie $W = -\vec{\mathfrak{M}} \cdot \vec{B} = -\mathfrak{M}_z B$. (dans la position normale, l'aimant est aligné avec le champ : c'est la position où l'énergie est la plus basse).



Une orbite d'électron est équivalente à un circuit fermé. De ce fait, un électron de moment cinétique $\vec{L}$ a un moment magnétique $\vec{\mathfrak{M}} = -\frac{e\hbar}{2mc}\frac{\vec{L}}{\hbar}$. $\frac{e\hbar}{2mc}$ est le magnéton de Bohr.



Les 3 possibilités d'orientation du moment cinétique $l = 1$ dans le champ $\vec{B}$. **Les trois énergies correspondantes.** **L'énergie en l'absence de champ magnétique.**

Résumons-nous : la valeur propre de l'énergie $E = -\frac{1}{4}$ rydberg (nombre quantique d'énergie $n = 2$) peut être obtenue de quatre façons différentes :

moment cinétique en unités $\hbar$	projection sur un axe arbitraire
0 (état *s*)	0 (1 fonction d'onde)
1 (état *p*)	+1, 0, −1 (3 fonctions d'onde possibles)

On dit que l'état $n = 2$ est « dégénéré » : lorsqu'on ne mesure que l'énergie, on ne connaît pas complètement l'état dans lequel se trouve l'atome, car son moment cinétique a plusieurs valeurs possibles.

Le même phénomène se produit pour les énergies supérieures. Ainsi pour $n = 3$, il y a un état *s* ($l = 0$), trois états *p* ($l = 1$) et cinq états *d* ($l = 2$).

Une comparaison peut aider à comprendre ce que sont ces états dégénérés. Les hommes nés en 1933 (la classe 1953) ont tous le même âge (trente et un ans en 1964). Ils sont tous dans un état propre de la grandeur « âge », correspondant à la même valeur propre, mais cet état est bien évidemment dégénéré (leur âge est fonction de leur seule date de naissance, et non du nombre ou du sexe de leurs enfants). Par exemple, parmi ces hommes qui ont tous le même âge, certains n'ont pas d'enfants, d'autres ont 1 enfant, d'autres 2, etc. On peut même remarquer que l'état « hommes à 2 enfants » est lui-même triplement dégénéré : ces 2 enfants peuvent être soit deux garçons, soit un garçon et une fille, soit deux filles.

On retrouvera les atomes en remplaçant :
l'âge par l'énergie;
le nombre d'enfants par le moment cinétique;
le nombre de garçons par la projection du moment cinétique.

Toutefois, il faut se garder de pousser trop loin la comparaison. Si nous voulons savoir combien André a de garçons (sachant qu'il a deux enfants), il suffit de le lui demander. La dégénérescence n'était qu'apparente et due à notre ignorance.
Si nous voulons savoir quelle est la projection L_z sur l'axe Oz du moment cinétique d'un atome dans l'état *p*, il faut interroger l'atome, c'est-à-dire mesurer L_z.

Pour cela, une méthode consiste à placer l'atome dans un champ magnétique parallèle à Oz : au point de vue magnétique, un électron en mouvement est équivalent à un petit aimant aligné sur son moment cinétique. Aux trois orientations possibles du moment cinétique par rapport à O_z, donc de l'aimant par rapport au champ magnétique, correspondent trois énergies différentes W, 0, et — W, et le niveau d'énergie E_2 se clive en trois niveaux : $E_2 + W$, E_2, et $E_2 - W$. (Dans le jargon de la mécanique quantique, on dit que « la dégénérescence est levée ».)

interprétation probabiliste

Il y a une différence fondamentale entre ce cas et celui de la réponse d'André à notre question sur le sexe de ses enfants : le fait d'interroger André n'a pas changé le sexe de ses enfants, lequel était déjà bien déterminé.

Dans le cas de l'atome dans l'état *p*, nous avons posé une question à laquelle l'atome a trois réponses possibles : $L_z = +\hbar$, $L_z = 0$, $L_z = -\hbar$, correspondant chacune à un état bien défini (état propre de L_z). **Or, avant de subir l'effet du champ magnétique, l'atome n'était pas nécessairement dans**

un de ces trois états. Il suffit de réfléchir un peu pour s'en convaincre; la direction Oz, qui joue au moment de la mesure un rôle privilégié, n'avait en général aucune raison d'être favorisée avant la mesure : l'atome pouvait tout aussi bien être dans un état propre de la projection du moment cinétique sur l'axe *x*, par exemple. C'est le fait de le placer dans un champ magnétique orienté suivant Oz qui l'a obligé à « sauter » de l'état où il se trouvait dans un des trois seuls états propres possibles de L_z!

Ainsi, mesurons la valeur de L_z de 3 000 atomes dans l'état *p*, préparés de façon identique (par échauffement, par exemple). Un résultat typique sera le suivant :

1 020 atomes donnent $L_z = +\hbar$;
970 atomes donnent $L_z = 0$;
1 010 atomes donnent $L_z = \hbar$.

En voici l'interprétation : **avant la mesure, nous ne savons rien de l'état où se trouve chaque atome; tout ce que nous avons appris, c'est que la probabilité d'obtenir, pour un atome pris au hasard, une des trois valeurs possibles de L_z est de 1/3 pour chacune de ces valeurs.**

En fait, cela n'est pas tellement choquant, car le même phénomène existe pour la lumière et est bien connu : la lumière d'une lampe à incandescence n'est pas polarisée.

Pour obtenir une lumière polarisée circulaire droite, par exemple, nous interposons un Polaroïd qui ne laisse passer que cette composante, ce qui réduit de moitié l'intensité de la lumière. L'interprétation quantique est : chaque photon a le choix entre la polarisation circulaire droite et la polarisation circulaire gauche, chacune avec une probabilité de 1/2.

avertissement pour le lecteur

Les considérations que nous avons développées en prenant pour exemple l'atome d'hydrogène et son moment cinétique se retrouvent dans l'étude des états propres de tout système à propos de bien d'autres grandeurs physiques. Elles peuvent être déduites de principes tout à fait généraux, qui sont exposés dans les quatre pages suivantes sur fond vert. Cet exposé, nécessairement abstrait, n'est pas absolument nécessaire à l'intelligence du texte, mais est essentiel pour quiconque veut vraiment comprendre l'esprit de la mécanique quantique.

quelques concepts nouveaux et révolutionnaires

Jusqu'ici, nous avons procédé par inductions successives : la particule a été « remplacée » par une onde, cette onde est devenue une onde de probabilité de présence ou « fonction d'onde », la fonction d'onde a finalement été définie comme étant la solution de l'équation de Schrödinger.

Moyennant quoi nous avons obtenu, pour des exemples simples, mais réels, une concordance absolument remarquable avec les faits expérimentaux : pénétration des barrières de potentiel, niveaux d'énergie de l'oscillateur harmonique ou de l'atome d'hydrogène.

Toutefois, dans le dernier cas, une certaine difficulté d'interprétation est apparue : à une même énergie correspondent plusieurs fonctions d'onde, c'est-à-dire plusieurs états.

Cela tient à ce que l'énergie n'est qu'une des grandeurs physiques qui définissent l'état d'un système. Tous les coureurs cyclistes de 75 kg faisant partie d'un même peloton ont la même énergie; ils ne sont pourtant pas identiques. Même un système aussi simple qu'un atome à un électron périphérique — une particule dans un champ central — n'est pas dans un état entièrement défini si l'on n'en connaît que l'énergie; effectivement, dans ce cas, nous avons constaté, pour une même énergie, la présence d'états différents, correspondant à certaines valeurs du moment cinétique et de sa projection sur un axe, valeurs elles-mêmes quantifiées.

Cela conduit à penser que, de même que l'équation de Schrödinger donne les énergies possibles d'un système, il doit y avoir des équations analogues pour les autres grandeurs physiques caractérisant les états possibles du système — Dirac a appelé ces grandeurs les **« observables ».**

les opérateurs

Procédons une dernière fois par induction, à partir de l'équation de Schrödinger (à une dimension pour simplifier, cela ne change rien au raisonnement) :

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx}+V\Psi=E\Psi.$$

Comparons à l'équation classique de conservation de l'énergie $\frac{p^2}{2m}+V=E$ (énergie cinétique + énergie potentielle = énergie totale constante).

Formellement, pour passer de l'une à l'autre, il suffit de remplacer l'expression classique de l'énergie cinétique par l'« opérateur » $-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}$, et de multiplier les deux membres par ψ.

Dans cette optique, l'équation de Schrödinger devient l'« équation aux valeurs propres de l'opérateur H » :

$$H=-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}+V.$$

Le mot « opérateur » se comprend presque de lui-même : un opérateur se définit par l'action — l'opération — qu'il exerce sur une fonction. La confusion naît de ce que, souvent — et par raison de simplicité —, l'opérateur se représente de la même façon qu'une expression mathématique simple.

Ainsi, 1 peut être soit le nombre 1, soit un opérateur qui redonne la fonction ψ identique à elle-même (c'est l'opération identité).

De même, x peut être tout simplement la variable x, mais aussi l'opérateur x, dont la signification évidente est de multiplier la fonction ψ par x : x opérant sur ψ = produit $x\psi$.

Tout naturellement, l'opérateur $\frac{d}{dx}$ appliqué à la fonction ψ commande la dérivation de cette fonction par rapport à x :

$$\frac{d}{dx}\cdot\psi=\frac{dx}{d\psi}.$$

valeurs propres et états propres d'un opérateur

De même que l'opérateur H de l'équation de Schrödinger, tout opérateur A a un ensemble de valeurs propres a, nombres tels que l'équation $A\psi = a\psi$ ait une solution ψ répondant à des conditions « aux limites », maintenant familières.

Jusqu'ici, à part une subtilité mathématique apparemment gratuite, rien que de très banal.

Ce qui l'est moins, c'est le principe physique que nous allons énoncer, et à partir duquel nous aurions pu dès le début bâtir la mécanique quantique, si nous n'avions craint d'assommer irrémédiablement le lecteur. Malgré son caractère abstrait, ce principe est la base du formalisme qui permet d'expliquer sans contradiction le monde des atomes et des particules. Le voici : **Toute grandeur physique se représente par un opérateur, et une mesure de cette grandeur physique ne peut donner comme résultat qu'une des valeurs propres de cet opérateur. Le système est alors dans un des états propres possibles pour cette grandeur physique.**

L'énergie d'un système en est évidemment le premier exemple — puisque c'est en partant de l'équation de Schrödinger que nous avons induit ce principe. La grandeur physique « énergie » se représente bien par un opérateur, et l'on trouve comme énergies possibles (niveaux des électrons atomiques ou niveaux d'une particule dans un puits de potentiel) les valeurs propres de cet opérateur.

Pour passer à d'autres applications, il suffit de savoir écrire les opérateurs correspondant aux diverses grandeurs physiques.

Pour les grandeurs classiques, pas de problème : elles sont soit fonction de x (et l'opérateur correspondant à x est la multiplication par x), soit fonction de la quantité de mouvement. Or, l'opérateur correspondant à $p^2{}_x$ est $-\hbar^2\frac{d^2}{dx^2}$ (d'après l'équation de Schrödinger), donc l'opérateur correspondant à

$$p_x \text{ est } -i\hbar\frac{d}{dx}\ (1).$$

Nous pouvons d'ailleurs immédiatement vérifier cette assertion en cherchant à résoudre l'équation

$$\underbrace{-i\hbar\frac{d}{dx}\psi}_{\text{opérateur}} = \underbrace{p}_{\text{nombre}} \times \underbrace{\psi}_{\text{(fonction propre)}},$$

qui doit nous donner les valeurs propres p_x de la quantité de mouvement.

La solution en est l'onde plane familière, $\psi = e^{i\frac{p_x}{\hbar}}$, et cette solution existe quelle que soit la quantité de mouvement p_x (on voit sur cet exemple que la mécanique quantique n'entraîne pas nécessairement la quantification!). Cela veut dire simplement que les états propres de la grandeur « quantité de mouvement » sont toutes les ondes planes possibles — résultat bien connu.

(1) On pourrait prendre aussi $+i\hbar\frac{d}{dx}$. Le choix n'a aucune importance, pourvu qu'on s'y tienne.

une illustration des principes quantiques : le moment cinétique

Le moment cinétique est une grandeur physique importante, car, au cours de l'évolution d'un système isolé, il reste constant au même titre que l'énergie.

Lorsque les particules n'ont pas de spin, comme dans le cas que nous étudions actuellement, cette proposition s'applique au moment cinétique classique, dit « orbital » (s'il y avait des particules à spin dans le système, c'est le moment cinétique total, somme du moment orbital et du spin, qui resterait constant). Mathématiquement, le moment cinétique, nous l'avons vu, est le produit (vectoriel) de la distance au centre par l'opérateur quantité de mouvement.

Du fait que c'est un vecteur, il y a trois opérateurs, correspondant chacun à une des trois projections sur les axes, et un opérateur, somme des carrés de ces trois opérateurs, pour le carré de la grandeur du moment cinétique, L^2.

D'après le principe énoncé ci-dessus, si l'on essaie de déterminer par une mesure la valeur d'une grandeur — L^2 ou l'une des trois projections, L_z par exemple —, on trouvera obligatoirement une des valeurs propres de l'opérateur correspondant, L^2 ou L_z.

La résolution, assez difficile, de l'équation aux valeurs propres pour L^2 donne comme seuls résultats possibles les valeurs $\hbar^2 l(l+1)$, où l est un entier; on peut donc se représenter le moment cinétique dans l'espace comme un vecteur de grandeur $\hbar\sqrt{l(l+1)}$. Cela n'est pas très loin du résultat de la théorie de Bohr ($\hbar l$); on dit, pour abréger, que le moment cinétique est l.

De plus, un état de moment cinétique l est $(2l+1)$ fois dégénéré, car l'équation $L^2\psi = \hbar^2 l(l+1)\psi$ a $2l+1$ solutions indépendantes. Ainsi l'état p $(l=1)$ est triplement dégénéré.

Reprenons notre analogie avec la situation familiale d'un homme marié : il peut avoir 0, 1, 2, ..., $2l$ enfants. Mais il y a plusieurs répartitions possibles des enfants entre garçons et filles (en fait $2l+1$). Par exemple, deux enfants peuvent être :

2 garçons	1 garçon	0 garçon
0 fille	1 fille	2 filles

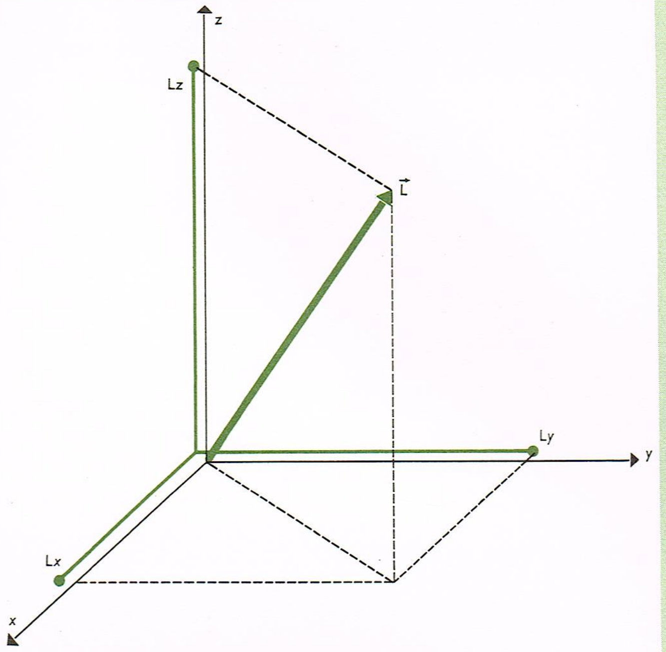
C'est un état triplement dégénéré! Dans le langage quantique, l'opérateur « nombre de garçons » appliqué à l'état « 2 enfants » a pour valeurs propres 0, 1 ou 2.

Si l'on cherche la valeur de la projection du moment cinétique sur l'axe z — ce qui correspond à chercher la projection du nombre d'enfants sur l'axe « nombre de garçons moins nombre de filles » —, il faut appliquer l'opérateur L_z. Si le système est dans l'état p, les valeurs propres de L_z sont ($m = -1, 0$ ou $+1$).

Cela signifie que si l'on effectue sur un atome de moment cinétique unité $(l=1)$ la mesure de la composante du moment cinétique suivant un axe, on trouvera $m = -1, 0$ ou $+1$ (plus généralement, au moment l peut correspondre toute valeur entière de m comprise entre $-l$ et $+l$).

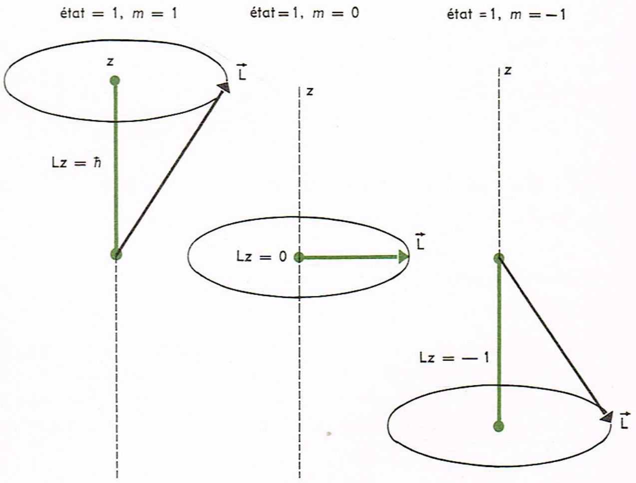
Ainsi que nous l'avons déjà souligné, cela ne veut pas dire qu'avant que soit faite une mesure qui donne par exemple $m = +1$ le système était dans cet état. Il n'était pas, en général, dans un état propre de L_z; **c'est la mesure qui a imposé à l'atome de se placer dans un état propre**; tant que rien d'autre n'agit sur lui, il va rester dans ce nouvel état.

Nous avons donc maintenant des atomes dont nous savons non seulement qu'ils sont dans l'état p $(l=1)$, mais aussi dans un état propre de L_z. Isolons, par exemple, les atomes avec $m=0$. Pouvons-nous connaître la valeur de leur moment cinétique suivant l'axe x? ($m=0$ nous dit seulement que $\vec{L}$ est dans le plan xy, mais $l=1$ et $m=0$ définissent complètement la fonction d'onde, et l'on constate que cette fonction n'est pas



Le moment cinétique $\vec{L}$ et ses composantes
$L^2 = L_x^2 + L_y^2 = L_z^2$.

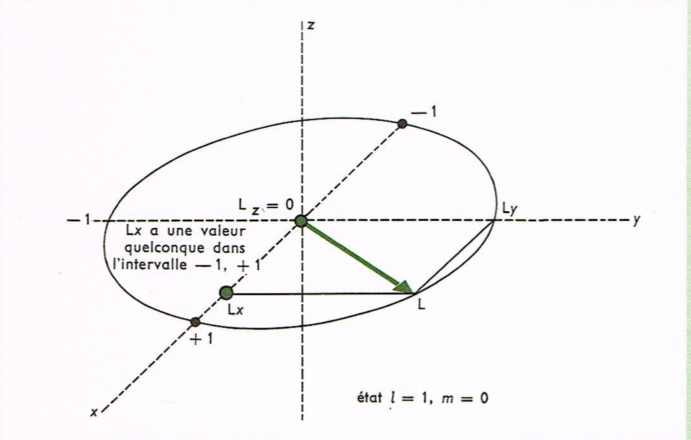
Expressions classiques	**Opérateurs**
$L_x = yP_z - zP_y$	$L_x = -i\hbar\left(y\frac{\delta}{\delta_z} - z\frac{\delta}{\delta_y}\right)$



Les trois états propres communs de L^2 et L_z dans l'état p.

Le vecteur $\vec{L}$ représente le moment cinétique dans l'espace : il a pour grandeur $\hbar\sqrt{l(l+1)}$, c'est-à-dire $\hbar\sqrt{2}$ pour l'état p, et pour projection L_z, constant si le système est dans un des états propres de L_z. On ne sait rien de plus sur le vecteur $\vec{L}$; tout ce qu'on peut faire, c'est se le « représenter » comme étant placé avec une égale probabilité sur n'importe quelle génératrice d'un cône de révolution autour de l'axe z, cône de hauteur L_z et d'apothème $(\vec{L})$.

Si le vecteur est dans un état propre $L_z = 0$, le vecteur L est indéterminé dans le plan xOy : le système n'est pas dans un état propre de L_z.



fonction propre de l'opérateur L_x). Donc, si l'on essaie de mesurer la valeur de L_x pour un ensemble de systèmes de même L_z connu, il n'en sortira pas une valeur unique, mais différentes valeurs propres de L_x, dans certaines proportions (calculables à l'avance, car la donnée préalable de L^2 et L_x détermine l'état du système). Mais, attention : **après cette mesure de L_x on ne saura plus quelle est la valeur de L_z !**

Ainsi, il est impossible de connaître simultanément deux projections du moment cinétique. Cette situation rappelle fortement le principe d'incertitude qui s'applique à toute tentative de détermination simultanée de la position et de la quantité de mouvement. De fait, ces deux phénomènes sont la conséquence d'une même relation mathématique entre les opérateurs x et p_x d'une part, L_x et L_z d'autre part — relation dite d'« anticommutation ». En revanche, on peut connaître simultanément le moment cinétique et sa projection sur un axe (ou encore les coordonnées x et y d'un point, ou les projections p_x et p_y de la quantité de mouvement) : on dit que ces opérateurs « commutent » (1).

petite digression mathématique, non sans rapport avec la physique

Les règles que nous venons d'illustrer par l'exemple du moment cinétique peuvent se généraliser ainsi :

A toute grandeur physique α correspond un opérateur A. Tout état d'un système pour lequel la grandeur α a une valeur bien définie a_i est représenté par une fonction ψ_i, qui est la solution de l'équation « aux valeurs propres » :

$$A\,\psi_i = a_i\,\psi_i.$$

ψ_i est une fonction des variables dont dépend A; dans l'espace défini par ces variables, la fonction ψ_i a la signification probabiliste que nous avons attribuée à la fonction d'onde d'une particule : **le carré de sa valeur absolue $|\psi_i|^2$ pour chaque ensemble de valeurs des variables donne la probabilité qu'ont les variables de prendre les valeurs correspondantes.**

Un exemple bien connu de fonctions propres est constitué par l'ensemble des fonctions trigonométriques qui représentent le fondamental et les harmoniques (sons purs) de la vibration d'une corde. Or, tout son tiré de cette corde peut s'écrire comme une somme de ces différents sons purs, chacun affecté d'une certaine amplitude (c'est la célèbre décomposition en série de Fourier, déjà vue).

Cette propriété mathématique s'étend aux fonctions propres de n'importe quel opérateur : toute fonction Φ peut se développer en série des fonctions propres ψ_i de l'opérateur A (2) :

$$\Phi = c_1\psi_1 + c_2\psi_2 + \ldots + c_n\psi_n + \ldots$$

Les coefficients c_i sont les amplitudes avec lesquelles les états propres de l'opérateur A sont représentés dans l'état décrit par la fonction Φ.

Par une coïncidence curieuse, l'arsenal mathématique ébauché ci-dessus avait été mis sous forme cohérente et immédiatement utilisable pour les physiciens dans le célèbre livre de Courant et Hilbert (1924) *les Méthodes mathématiques de la physique*, publié juste un an avant que Schrödinger écrivît son équation.

Revenons à la physique. A propos d'un ensemble d'atomes identiques, c'est-à-dire tous dans le même état décrit par la même fonction ψ, posons-nous la question : quel résultat va donner la mesure de la grandeur α? Si ψ est une fonction propre ψ_i de l'opérateur A, correspondant à la valeur propre a_i, pas de problème : la mesure donnera toujours — par définition — a_i.

Si ψ n'est pas fonction propre de A, nous savons que la mesure est susceptible de donner les différentes valeurs propres de A — et ces valeurs seulement. Le développement ci-dessus permet de calculer la probabilité avec laquelle chaque valeur propre va figurer. En effet, la valeur propre a_k est obtenue lorsque le système est dans l'état ψ_k; or, l'état ψ_k intervient dans le développement avec l'amplitude c_k, donc la probabilité $|c_k|^2$. (Tout se passe comme si le système était dans chacun des états propres pendant une fraction de temps proportionnelle à l'intensité avec laquelle chaque fonction propre est représentée dans la fonction d'onde.)

Par exemple, la fonction d'onde décrivant l'état d'une particule entre deux murs placés en $x = 0$, $x = a$, est $\psi_n = \sin\left(\frac{n\pi x}{a}\right)$, ce qui peut s'écrire encore : $\psi_n = \frac{1}{2i}\left[e^{in\pi x/a} - e^{-in\pi x/a}\right]$.

Sous cette forme, on voit que ψ_n est équivalent à la superposition de deux ondes planes, c'est-à-dire de deux états propres de l'opérateur quantité de mouvement, avec les valeurs propres

$$p_1 = \frac{n\pi x}{a} \quad \text{et} \quad p_2 = \frac{n\pi x}{a}.$$

Si donc on mesure la quantité de mouvement de la particule, on trouvera soit la valeur p_1, soit la valeur p_2 — avec des probabilités égales, ce qui, pour une fois, correspond bien au sens commun.

---

(1) Ces mots un peu rébarbatifs de « commutation » et d' « anticommutation » sont introduits ici uniquement pour familiariser le lecteur avec des termes qu'il retrouvera constamment si la fantaisie lui prend d'approfondir sa connaissance de la mécanique quantique.

(2) Cela provient du fait que les fonctions ψ forment un « ensemble complet ».

qu'est-ce que le spin?

Au point de vue de l'expérimentateur, le spin de l'électron se présente comme un moment cinétique attaché à l'électron (moment cinétique « intrinsèque »), et dont la projection sur un axe peut prendre l'une ou l'autre des deux valeurs $\frac{\hbar}{2}$ et $-\frac{\hbar}{2}$.

Par exemple, si l'on fait passer dans le dispositif de Stern et Gerlach un faisceau d'atomes à un électron périphérique (atomes alcalins : lithium, sodium, etc.), on obtiendra deux taches. L'une correspond au spin de l'électron suivant la direction du champ magnétique, l'autre au spin opposé à la direction du champ magnétique.

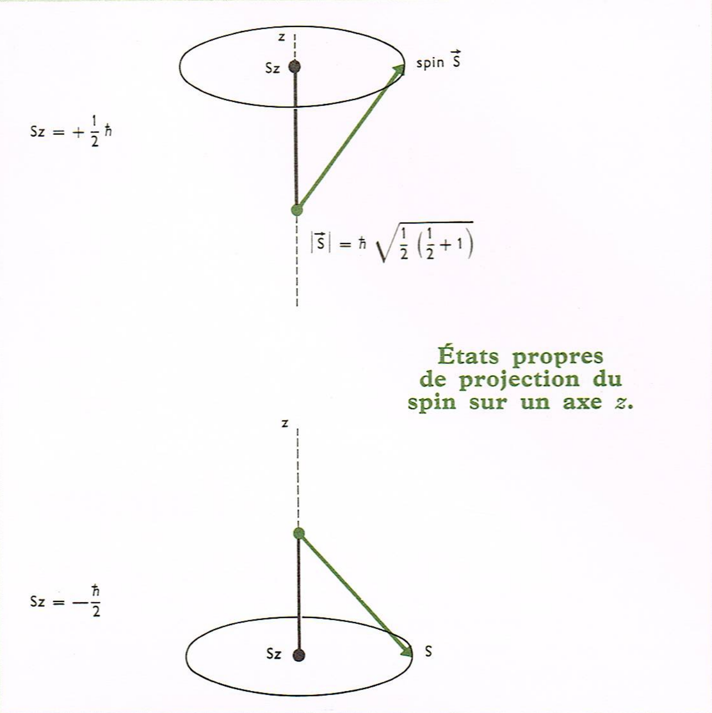
De plus, dans le cas le plus fréquent où il n'existe avant la mesure aucune direction privilégiée dans l'espace (faisceau « non polarisé »), les deux taches sont d'intensité égale — chaque atome a une probabilité égale à 1/2 d'aligner son spin suivant le champ, ou dans un sens contraire à lui.

On peut donc décrire l'état où se trouvent les atomes, avant la mesure, comme une combinaison linéaire, à poids égaux, des deux états « projection du spin $= \hbar/2$ » et « projection du spin $= -\hbar/2$ ».

Le dédoublement de la raie D du sodium (raie émise dans la transition de l'état $2p$ à l'état $1s$) s'explique aussi très simplement en tenant compte du spin de l'électron. Dans l'état p (moment cinétique orbital unité), le spin de l'électron peut soit s'aligner avec le moment cinétique orbital, soit s'y opposer, ce qui correspond à deux états d'énergies légèrement différentes (structure « fine »).

En physique atomique, tous les problèmes où intervient le spin peuvent se traiter en considérant le spin comme un moment cinétique, sujet aux mêmes règles que le moment cinétique ordinaire (« orbital »), avec ces particularités que son nombre quantique est $s = \frac{1}{2}$ (et non $l = 0, 1, 2, \ldots$) et que son moment magnétique est « anormal » (le moment magnétique lié au spin de l'électron est égal à un magnéton de Bohr — la même valeur que pour un moment cinétique orbital $l = 1$).

La mécanique quantique, dans sa forme relativiste, fait encore mieux : l'équation de Dirac — équation relativiste applicable à l'électron — introduit automatiquement le spin de l'électron et rend compte de la valeur de son moment magnétique. De plus, cette équation permet de prévoir l'existence d'un « anti-électron » : c'est l'électron de charge positive, ou positon, découvert effectivement quelques années plus tard (Anderson, 1932).



États propres de projection du spin sur un axe *z*.

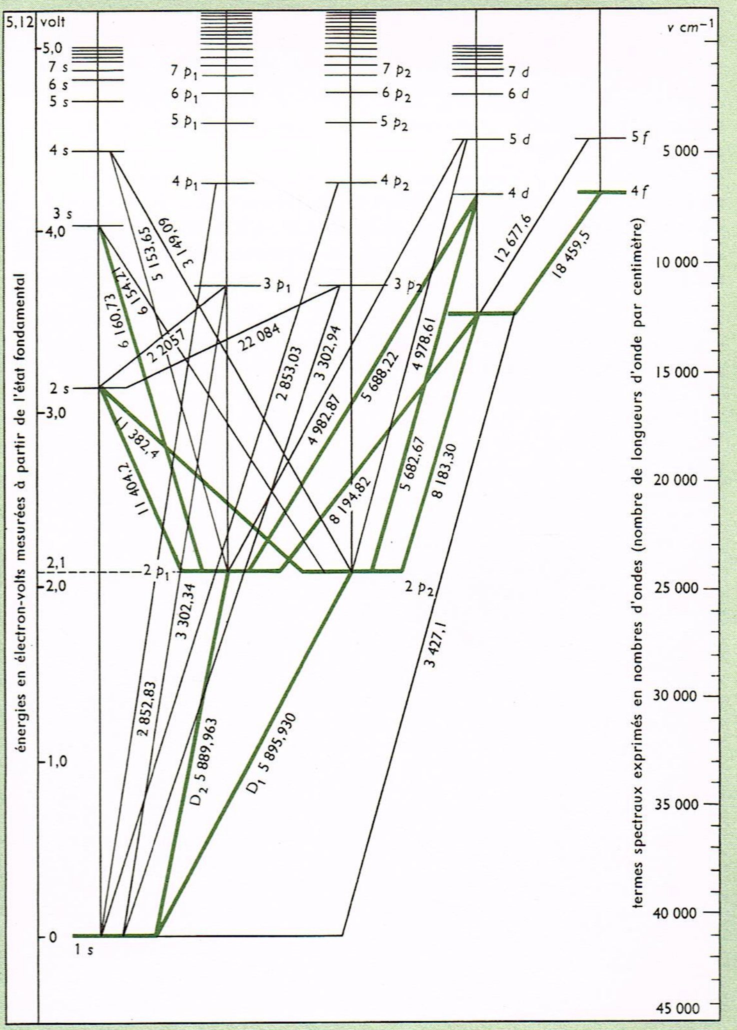


Diagramme des niveaux d'énergie de l'atome de sodium.

Ce diagramme représente les différents états de l'atome de sodium. Chaque état est schématisé par un trait horizontal dont l'ordonnée est proportionnelle à l'énergie (le zéro des énergies est ici l'état fondamental (*s*). L'étiquetage des états est le suivant :

le premier nombre est le nombre quantique d'énergie ; la lettre qui suit symbolise le moment cinétique orbital ; l'indice 1 ou 2 signifie que le spin est parallèle ou opposé au moment cinétique orbital.

A remarquer : la différence d'énergie entre l'état 2 *s* et l'état 2 *p* (la dégénérescence entre ces états n'existe que pour l'atome d'hydrogène, à cause du potentiel en 1/*r*), le dédoublement de l'état *p* et la faible valeur de la différence entre ces deux états (d'où le nom de structure fine). Le même dédoublement existe pour les états *d* et *f*, mais n'est pas représenté.

deux systèmes simples et deux phénomènes nouveaux

Après l'atome d'hydrogène et ses confrères les atomes alcalins (à un électron périphérique évoluant dans un champ central), passons aux systèmes qui les suivent dans l'échelle de complexité croissante : l'atome d'hélium (atome à deux électrons) et la molécule d'hydrogène (assemblage stable de deux atomes d'hydrogène). Cette étude va nous faire découvrir de nouveaux phénomènes, purement quantiques. Armés de ces outils, et forts de la théorie des atomes et des molécules, nous pourrons ensuite essayer de comprendre les édifices formés avec ces briques — c'est-à-dire les différents aspects de la matière telle qu'elle se présente à nos sens, solide, liquide ou gazeuse. Ce sera le but du prochain chapitre. Terminons-en donc d'abord avec les systèmes les plus simples.

l'atome d'hélium

L'hélium est l'élément le plus simple après l'hydrogène. Son noyau, composé de deux protons et deux neutrons, a une charge positive double : il faut donc deux électrons pour constituer un atome neutre. Nous sommes en présence de l'analogue quantique du célèbre problème dit « à trois corps » de la mécanique classique, problème dont il est impossible de donner une solution rigoureuse. Pour simplifier, parlons le langage de Bohr : les deux électrons décrivent chacun une orbite — comme la Terre et Mars autour du Soleil. Si chaque électron était seul, il décrirait une des orbites de Bohr de l'atome à un électron. La présence de l'autre perturbe, c'est-à-dire déforme, cette orbite — comme la présence de Mars déforme l'orbite de la Terre.

Mais, alors que la perturbation du mouvement de la Terre par Mars est petite, car la masse de Mars est beaucoup plus petite que celle du Soleil, la perturbation du mouvement d'un électron par l'autre est presque aussi importante que l'action du noyau sur l'électron : en effet, la charge d'un électron est la moitié de celle du noyau, et la distance électron-électron est en moyenne du même ordre que la distance électron-noyau.

Dans ces conditions, nous serons donc satisfaits si une méthode approchée nous donne des résultats numériques assez proches des valeurs expérimentales. Cette méthode est la suivante : on utilise la fonction d'onde qui représenterait le système s'il n'y avait pas de perturbation (c'est-à-dire, dans le cas de l'atome d'hélium, s'il n'y avait pas de répulsion électrostatique entre les deux électrons) et on calcule l'effet de cette perturbation sur l'énergie du système — en supposant que la fonction d'onde est, elle, peu modifiée par la perturbation.

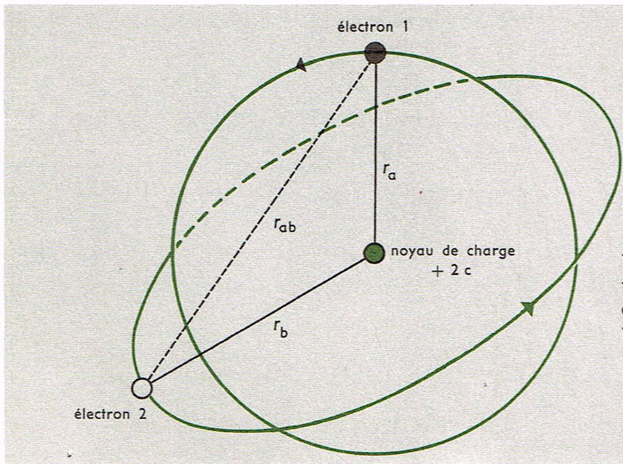
Ainsi, l'état fondamental se représente en écrivant que chacun des deux électrons, non influencé par l'autre, est dans l'état le plus bas — c'est-à-dire l'état 1*s* — du système noyau-électron unique. L'énergie de liaison est alors la somme des énergies de liaison (négatives) de chacun des deux électrons dans l'état *s*, et de l'énergie (positive) résultant de la répulsion des deux électrons. Cette répulsion est déterminée par le potentiel électrostatique familier :

$$V = \frac{e^2}{r_{ab}}.$$

L'énergie correspondante est la valeur moyenne de ce potentiel V, c'est-à-dire la valeur que l'on obtient en pondérant les diverses valeurs que prend V suivant les configurations électroniques par la probabilité de chaque configuration, c'est-à-dire $|\psi|^2$.

L'approximation consiste à prendre pour ψ le produit des fonctions d'onde qui décriraient l'état de chaque électron en l'absence de l'autre : la valeur calculée pour l'énergie de liaison de l'état fondamental, en employant cette approximation, est proche de la valeur expérimentale. Des méthodes mathématiques plus raffinées rendent l'accord presque parfait — à moins d'un millionième près.

Notre description de l'atome d'hélium est donc fondamentalement correcte. De la confrontation de cette description et des données expérimentales va résulter un nouveau principe, d'une importance extrême.



Une représentation planétaire possible d'un atome d'hélium : deux électrons, tournant chacun sur son orbite autour du noyau.

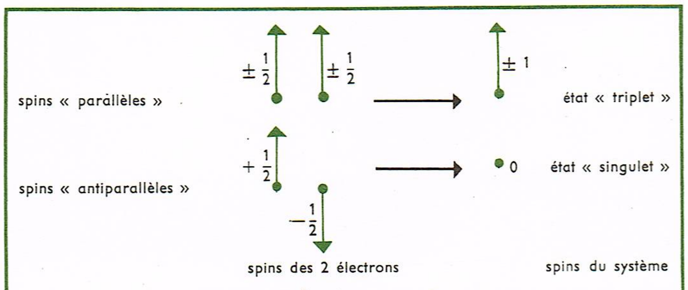


Wolfgang Pauli.

Phot. Keystone.

le célèbre principe d'exclusion de Pauli

Dans l'état fondamental de l'atome d'hélium, chacun des deux électrons a un moment cinétique « orbital » nul. Le moment cinétique résultant est donc lui aussi nul; mais l'atome peut avoir un moment cinétique résultant de la combinaison des spins des deux électrons. Pour simplifier le langage, représentons le spin $\vec{S}$ comme s'il était aligné avec sa projection S_z, et exprimons comme d'habitude tous les moments cinétiques en prenant $\hbar$ comme unité.



Combinaison des spins de 2 électrons

Les deux spins peuvent se combiner, pour donner un moment cinétique 1 s'ils sont parallèles $\left(\frac{1}{2}+\frac{1}{2}=1\right)$, 0 s'ils sont « antiparallèles » $\left(\frac{1}{2}-\frac{1}{2}=0\right)$.

Or, c'est un fait expérimental que, dans l'état fondamental, l'atome d'hélium a un moment cinétique nul (par exemple, un jet d'atomes d'hélium donne une seule tache dans l'expérience de Stern et Gerlach, et non les trois taches qu'il devrait donner si le moment cinétique était unité). Donc les deux électrons, qui sont chacun dans le même état d'énergie et de moment cinétique (orbital), c'est-à-dire sont décrits par les mêmes nombres quantiques d'espace ($n = 0$, $l = 0$, $m = 0$ pour l'état *s*), ont des nombres quantiques de spin différents : $S_z = +1/2$ et $S_z = -1/2$.

L'examen de l'ensemble des termes spectraux de l'atome d'hélium révèle que ceci est une loi générale : **deux électrons dans le même état orbital ne sont jamais dans le même état de spin; ou encore : un état d'espace et de spins, caractérisé par les quatre nombres *n*, *l*, *m*, s_z ne peut être occupé que par un seul électron.**

C'est là le principe d'exclusion, dû à Pauli, principe à partir duquel on peut, comme il a été fait précédemment, reconstruire la table des éléments (table de Mendeleïev). Ce principe joue aussi un rôle extrêmement important dans les systèmes formés d'un grand nombre de particules à spin 1/2 (nucléons des noyaux, électrons libres des métaux), où il empêche les particules de se « tasser » toutes dans l'état d'énergie le plus bas possible. C'est la base de la statistique de Fermi-Dirac.

Il existe une autre statistique quantique, dite de « Bose-Einstein », applicable aux particules de spin entier (photons, atomes d'hélium, etc.) pour lesquelles le principe d'exclusion ne joue pas. Le chapitre suivant présentera des exemples qui illustreront la différence entre ces deux statistiques.

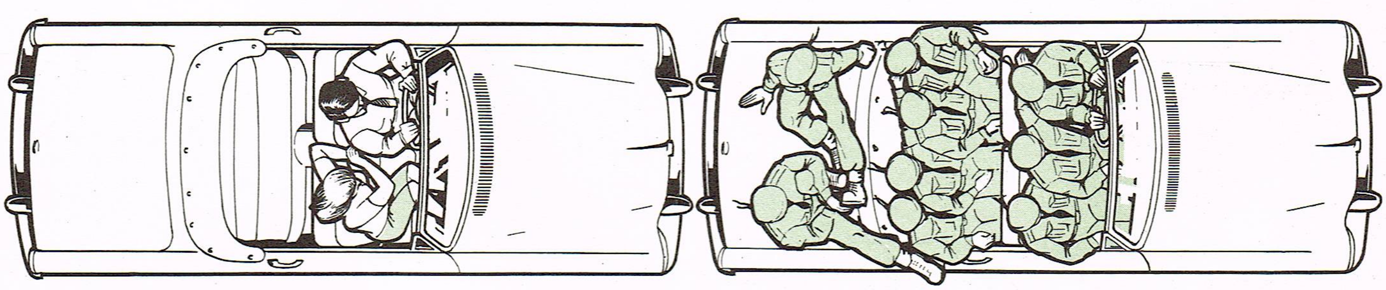
En attendant, on peut caractériser la statistique de Fermi-Dirac ainsi : elle s'applique à des particules qui peuvent exister sous deux états. Chaque particule occupe une « case » dans l'espace (des phases) et ne peut partager cette case qu'avec une particule dans l'autre état. Ainsi, dans la vie normale, une case (appartement ou villa) est occupée par un couple homme-femme, et un seul. Dans la statistique de Bose-Einstein, au contraire, une case (de l'espace des phases) peut être occupée par autant de particules que l'on veut (on passe de l'appartement pour ménage au dortoir de caserne, ce qui est évidemment beaucoup plus économique au point de vue espace).

A la différence des humains, qui, suivant les circonstances, peuvent vivre par couples ou en dortoirs, les particules n'ont pas le choix :

— les particules de spin 1/2 (électrons, protons, neutrons, etc.) obéissent à la statistique de Fermi-Dirac;

— les particules de spin entier (photons, mésons, atomes d'hélium, noyaux de deutérium, etc.) obéissent à la statistique de Bose-Einstein.

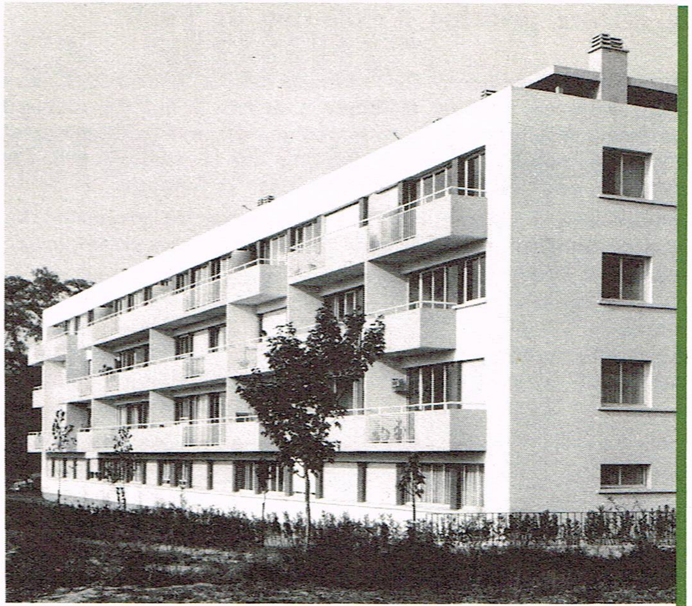
Deux exemples illustrant le principe de Pauli.



La présence d'un homme et d'une femme exclut la présence d'une troisième personne, laquelle serait nécessairement du même sexe que l'un ou l'autre (*three is a crowd...*). L'homme et la femme représentent les deux états d'une particule de Fermi-Dirac (électron, proton, neutron, toutes particules à deux états de spin), tandis que les soldats représentent les particules de Bose-Einstein (photons, mésons π, noyaux d'hélium, etc.). Du fait qu'une case de l'espace peut être occupée au plus par deux fermions identiques, mais par autant de bosons que l'on veut, la statistique de Fermi est très différente de la statistique de Bose. Un gaz d'électrons ne se laisse pas comprimer aussi facilement qu'un gaz de photons.

(V. la suite p. 304.)





Cette école *(à gauche)* et l'immeuble H. L. M. occupent à peu près le même volume. Pourtant, l'école abrite 300 élèves, alors que l'H. L. M. abrite à peine une centaine d'habitants. L'école est occupée par des bosons (les élèves, 30 dans une classe, réfectoire commun), l'H. L. M. par des fermions (un homme et une femme, parfois des enfants, dans chaque appartement).

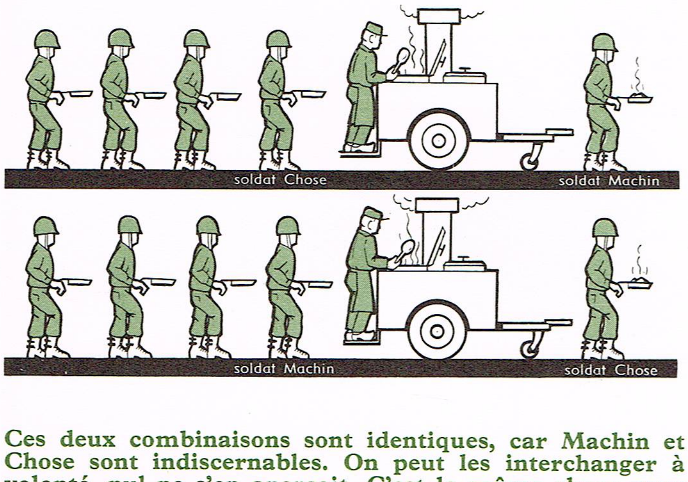
l'indiscernabilité des particules élémentaires

Les statistiques quantiques, Fermi-Dirac ou Bose-Einstein, ont entre elles un point commun qui les oppose à la statistique classique de Boltzmann : elles postulent l'indiscernabilité de deux particules identiques. En termes plus simples, rien ne distingue la situation où l'électron 1 est en *a* et l'électron 2 en *b*, de celle où l'électron 2 est en *a* et l'électron 1 en *b*. En fait, la phrase précédente n'a même aucun sens : **on ne peut dire quel électron est en *a*, quel autre est en *b*; tout ce que l'on peut dire, c'est que l'un des deux est en *a*, l'autre en *b*.**

Ces deux situations n'en constituent qu'une. Ou encore, pour parler le langage quantique, elles font partie du même état du système et ont la même probabilité. Donc la fonction ψ (*a*, *b*,), dont le carré donne la probabilité que l'électron 1 soit en *a*, l'électron 2 en *b*, doit donner la même probabilité pour 1 en *b*, 2 en *a*. On en déduit facilement que la fonction ψ a la propriété suivante : $\psi(a, b) = \pm \psi(b, a)$.

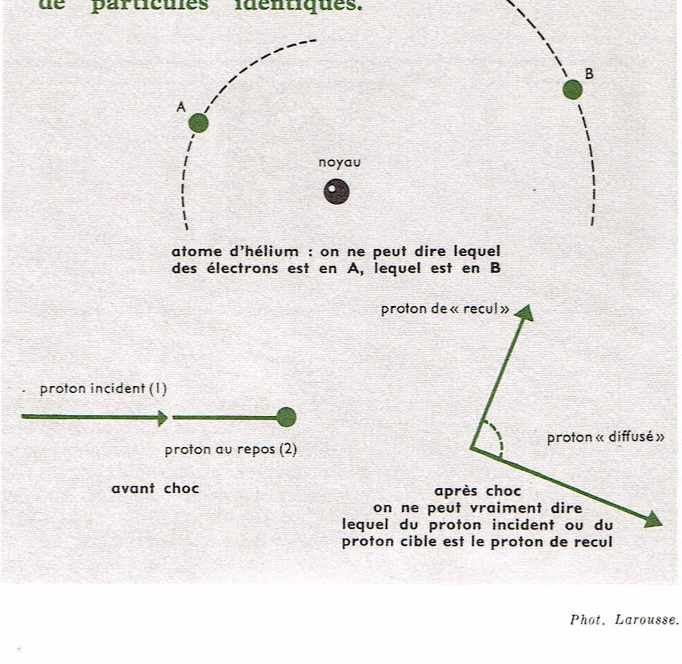
Lorsqu'on échange les positions de deux particules identiques, la fonction d'onde soit se retrouve identique, soit est multipliée par — 1; dans le premier cas, elle est dite « symétrique », dans le second « antisymétrique ».

Dans l'exemple simple de l'atome d'hélium dans son état fondamental, la fonction d'onde d'espace de l'ensemble des deux



Ces deux combinaisons sont identiques, car Machin et Chose sont indiscernables. On peut les interchanger à volonté, nul ne s'en aperçoit. C'est la même chose pour des particules identiques (photons du rayonnement du corps noir, électrons d'un atome, protons d'un noyau) : si on les échange, le système reste le même. Cela réduit considérablement le nombre de combinaisons possibles : c'est une première différence entre les statistiques quantiques et la statistique classique de Boltzmann.

Exemple d'indiscernabilité de particules identiques.



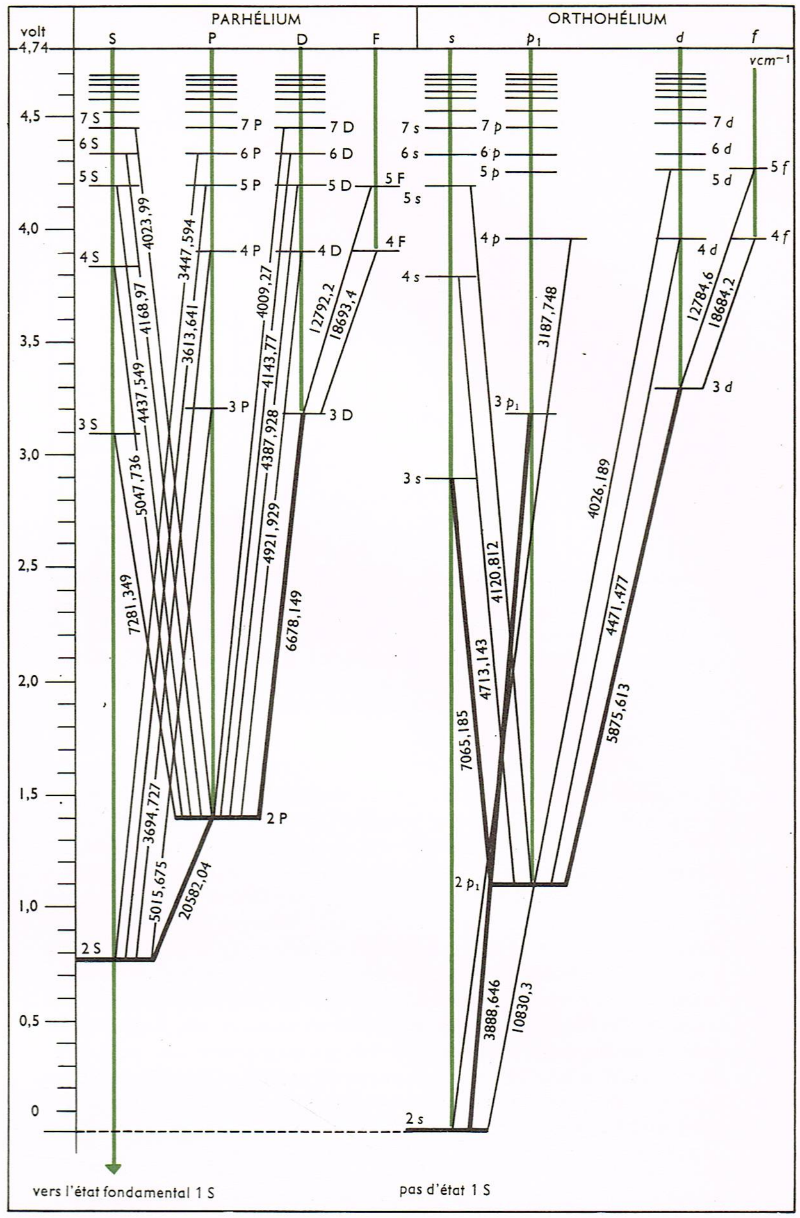
Phot. Larousse.

électrons est le produit des fonctions représentant chaque électron dans l'état 1s, fonctions que nous symboliserons par la lettre s : $\psi(a, b) = s(a).\,s(b)$. Elle est de toute évidence symétrique.

La fonction d'onde totale du système est le produit de cette fonction d'espace par la fonction décrivant l'état de spin des deux électrons. Expérimentalement, nous l'avons vu, seul l'état à spins antiparallèles, dit « état singulet », existe. Or, la fonction de spin correspondante est antisymétrique (alors que la fonction de spin de l'état à spins parallèles, dit « état triplet », est symétrique). Donc, la fonction d'onde totale du système est « antisymétrique ». Ainsi, pour les électrons, et plus généralement pour toutes les particules de spin 1/2, la fonction d'onde change de signe quand on échange toutes les coordonnées, de spin et d'espace, de deux particules identiques.

C'est une autre formulation du principe de Pauli, formulation qui s'appuie sur une loi physique extrêmement générale : un système de particules de spin 1/2 a une fonction d'onde antisymétrique (et répond à la statistique de Fermi-Dirac), alors qu'un système de particules de spin entier a une fonction d'onde symétrique (et répond à la statistique de Bose-Einstein).

Cette relation extrêmement générale entre spin et statistique n'a été démontrée que tout récemment par Pauli, peu de temps avant sa mort.



Raies spectrales de l'atome d'hélium.

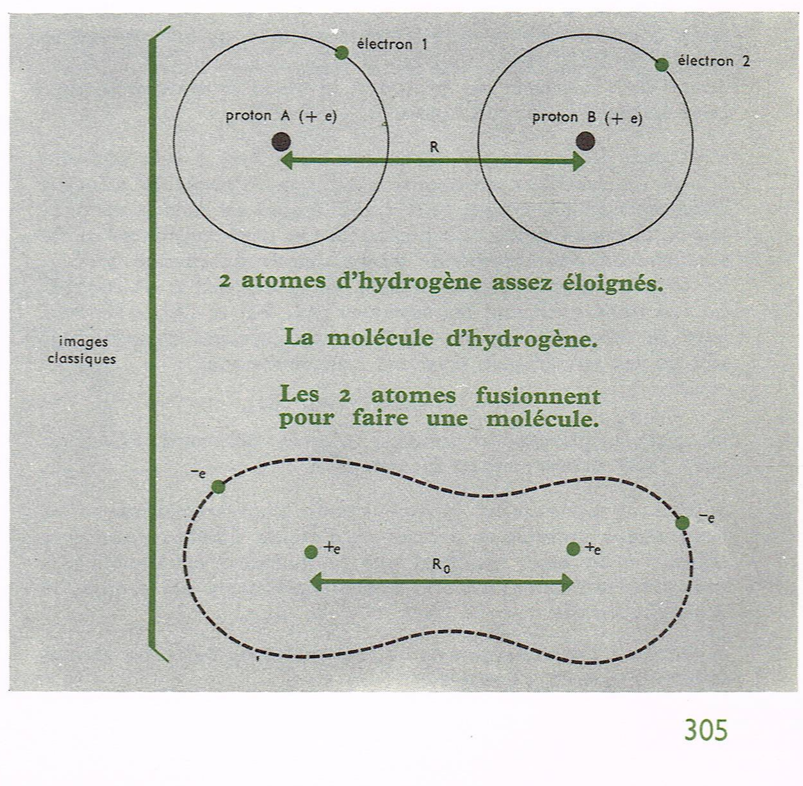
La moitié de gauche représente les états symétriques d'espace (parhélium : états singulets, à spins antiparallèles, antisymétriques de spin), la moitié de droite représente les états antisymétriques d'espace (orthohélium : états triplets, à spins parallèles, symétriques de spin).
Il n'y a aucune transition entre ces deux séries d'états, et l'état fondamental n'existe pas dans la colonne de droite.

la molécule d'hydrogène

Ayant maintenant compris l'essentiel des principes qui entrent en jeu dans la constitution des atomes, nous pouvons passer à l'étape suivante : **comment les atomes s'assemblent-ils pour former des molécules?**

On comprend sans grande difficulté la liaison dite « polaire », illustrée par la combinaison du chlore et du sodium pour former le sel marin. L'atome de chlore possède sept électrons périphériques et complète sa couche à huit en prenant l'électron périphérique du sodium; il devient ainsi un ion chlore négatif, alors que le sodium est maintenant un ion positif, et l'attraction électrostatique les lie ensemble pour former la molécule de chlorure de sodium.
Ce type de liaison est très fréquent dans les réactions de la chimie minérale, mais ne peut expliquer la liaison de deux atomes identiques. Comment deux atomes d'hydrogène s'associent-ils pour former la molécule d'hydrogène? La physique classique a buté sur ce problème pendant un siècle; aussi sa résolution en 1927 par Heitler et London est-elle considérée comme un des grands succès de la mécanique quantique.

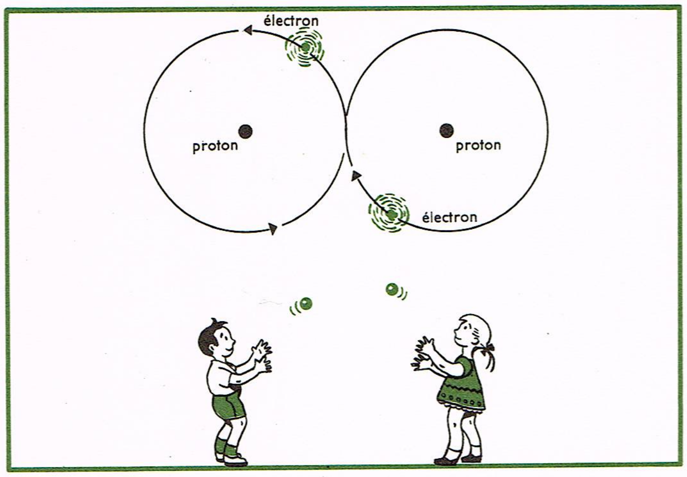
Que nous dit l'expérience? Que deux atomes s'unissent pour former une molécule et que cette molécule n'a pas tendance à



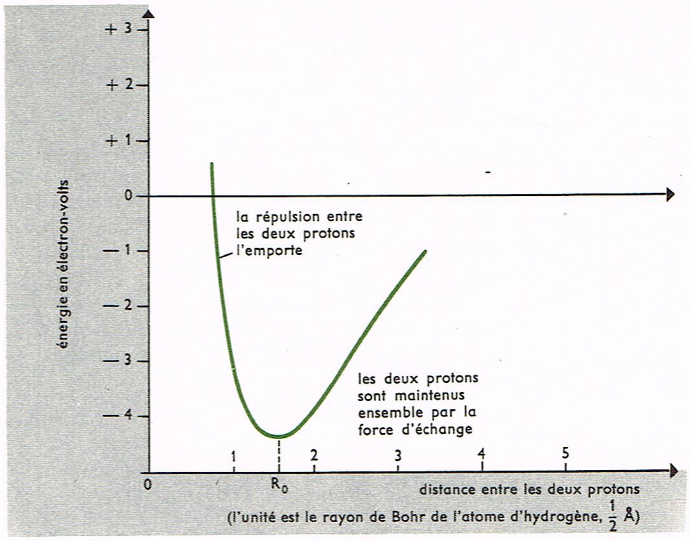
2 atomes d'hydrogène assez éloignés.

La molécule d'hydrogène.

Les 2 atomes fusionnent pour faire une molécule.



Les deux protons échangent continuellement leurs électrons — comme deux enfants échangent une balle. Si les enfants ne jouaient pas à la balle, ils iraient jouer chacun de leur côté : c'est l'échange de la balle qui les maintient ensemble.



Représentation de l'attraction entre les deux atomes d'hydrogène en fonction de la distance de leurs centres.

se scinder d'elle-même en deux atomes. C'est donc que son énergie de liaison est plus grande (en valeur absolue) que la somme des énergies de liaison des deux atomes. On peut représenter le phénomène de la façon suivante : deux atomes initialement très éloignés l'un de l'autre ont tendance à se lier quand ils se rapprochent, et, par suite, leur énergie totale diminue. Mais, au-dessous d'une certaine distance R_0 entre les deux protons, la répulsion électrostatique entre les deux protons l'emporte sur la liaison due à la couche électronique. La distance R_0 est une distance d'équilibre, autour de laquelle les deux protons vont exécuter de petites vibrations (au voisinage de ce point d'équilibre, la courbe d'énergie a la même allure que celle de l'oscillateur harmonique, vue page 291). Ces vibrations étant petites à cause de la masse élevée des protons, nous raisonnerons maintenant comme si R_0 était fixe.

Reste à comprendre comment les deux protons à la distance R_0 sont liés par leurs électrons. Appliquons la même méthode que pour l'atome d'hélium. L'état le plus bas en énergie du système constitué par deux atomes, très distants, est représenté par le produit de la fonction d'onde décrivant un électron dans l'état 1s autour du proton A et de la fonction analogue pour le second électron autour du proton B.

Mais cette forme simple ne convient pas tout à fait, car elle différencie les deux électrons, alors que, d'après le principe d'indiscernabilité, on ne peut savoir lequel est autour du proton A, et lequel autour du proton B. On peut considérer qu'ils sont chacun pour moitié du temps autour de chaque proton. La fonction d'onde s'obtient donc en combinant à poids égal les fonctions décrivant les systèmes (A1, B2) et (A2, B1); cela peut se faire soit en les additionnant (système symétrique), soit en les soustrayant (système antisymétrique) :

$$\psi = (A1)\ (B2) \pm (A2)\ (B1),$$

où (A1) symbolise la fonction d'onde décrivant l'électron dans l'état ls par rapport au proton A.

Supposons donc (c'est là que réside l'approximation, mais nous avons constaté à propos de l'atome d'hélium que cela conduit à de bons résultats) que la molécule est décrite par une telle fonction d'onde. L'énergie est alors la somme de plusieurs termes :

— la somme des énergies de liaison de chacun des deux atomes (c'est notre zéro d'énergie);

— l'énergie due à la répulsion entre les deux électrons, et entre les deux protons;

— enfin, l'énergie due à l'attraction du proton A sur l'électron lié au proton B et du proton B sur l'électron lié à A.

Seul ce dernier terme peut expliquer la liaison de la molécule. Encore faut-il pour cela qu'il soit plus grand (en valeur absolue) que l'énergie de répulsion proton-proton et électron-électron.

Or, sa valeur est donnée par la moyenne du potentiel correspondant, moyenne que l'on obtient, comme toujours, en pondérant par le carré de la fonction d'onde. On voit facilement, en procédant à cette élévation au carré, que l'énergie d'attraction proton-électron est la somme de deux termes :

- dans le premier, le potentiel coulombien e^2/r_{2A} est multiplié par le carré de la fonction d'onde (2B). C'est l'énergie due à l'attraction par le proton A de l'électron lié au proton B, terme prévu dans le modèle classique. Le calcul montre qu'il est trop petit pour expliquer la liaison — c'est bien là-dessus qu'a buté pendant longtemps toute tentative d'explication de la molécule d'hydrogène;
- dans le second, l'habituel carré de fonction d'onde est remplacé pour chaque électron par le produit de deux fonctions d'onde, l'une le décrivant comme lié au proton A, l'autre comme lié au proton B. Une telle expression n'a aucun analogue classique : c'est la traduction mathématique de l'interprétation initiale à laquelle nous a contraints le principe d'indiscernabilité, à savoir que chaque électron est lié tantôt à l'un des protons, tantôt à l'autre, ou encore qu'il y a échange continuel des électrons entre les deux protons. **D'où le nom d' « énergie d'échange » pour ce terme nouveau, purement quantique.** La fonction d'onde symétrique donne une énergie d'échange négative, qui assure la liaison de la molécule.

L'importance du phénomène d'échange apparaît clairement; c'est à lui que l'on doit la liaison de la molécule d'hydrogène (puisque le terme classique est insuffisant pour l'assurer).

Jusqu'ici nous n'avons parlé que de la partie de la fonction d'onde qui fournit la probabilité de présence dans l'espace. L'orientation relative des spins des deux électrons s'en déduit immédiatement, puisque la fonction d'onde totale doit être antisymétrique. Ainsi, dans la molécule d'hydrogène, la partie

spatiale étant symétrique, la fonction de spin doit être antisymétrique. Les spins des deux électrons sont donc antiparallèles, et, par suite, la molécule d'hydrogène a un moment cinétique total nul. Ce que confirme l'expérience.

le rôle fondamental des forces d'échange

C'est ainsi, sans aucun artifice, qu'un des problèmes fondamentaux de la chimie fut résolu — deux ans après la naissance de la mécanique quantique —, et, à sa suite, le mystère de la « saturation » des liaisons chimiques. En effet, pourquoi n'y a-t-il pas de molécule d'hydrogène à trois atomes? Tout simplement parce que, à cause du principe d'exclusion, l'électron du troisième atome ne pourrait s'échanger qu'avec celui des électrons de la molécule qui a même direction de spin — d'où répulsion!

Ce phénomène de saturation se retrouve dans la constitution des noyaux; que le noyau soit formé de 4 nucléons (hélium) ou de 238 (uranium), l'énergie de liaison par nucléon est la même, exactement comme si chaque nucléon n'était lié de façon vraiment forte qu'à ses « voisins ». Tout naturellement, la même explication par les forces d'échange est venue à l'esprit : c'est la base de la théorie — encore imparfaite, il faut bien le dire — des forces nucléaires; dans le noyau de deutérium, par exemple, le proton et le neutron sont liés par échange continuel de « mésons ».

une science en éternel devenir

Nous voilà parvenus au terme de cette étude de la mécanique quantique et de ses applications aux phénomènes atomiques et moléculaires. Répétons-le, si ces pages ont pu paraître difficiles, elles ne constituent qu'un préambule à la physique moderne.

Le chapitre suivant montrera comment, à partir des édifices élémentaires, on explique l'aspect et les propriétés de la matière en général et des solides en particulier; cette « physique des solides », qui, en principe, semblerait n'être qu'application, est en fait extrêmement compliquée, mais son évolution reste encore pleinement prometteuse, malgré les succès qu'elle a déjà remportés.

Dans le domaine de l'infiniment petit, les succès initiaux de la mécanique quantique n'ont pas, et il s'en faut, liquidé la question, mais les concepts d'état et de valeur propres et l'interprétation probabiliste tiennent toujours; **mémorables années que celles où Bohr, de Broglie, Heisenberg, Schrödinger, Born, Dirac, Pauli, etc., bouleversèrent la pensée scientifique!**

Depuis, bien des problèmes ont surgi, de façon très naturelle : souvent, la bonne concordance entre résultats expérimentaux et prévisions théoriques n'est qu'approximative. Une expérience ultérieure peut révéler une légère discordance ou un fait nouveau, qui obligent les théoriciens à repenser le problème — processus très fécond, qui a amené, par exemple, Lee et Yang à douter de la « conservation de la parité » (1956) [1]. Un autre exemple célèbre est fourni par l'effet Lamb : en utilisant les techniques les plus modernes d'hyperfréquences, Lamb découvrit en 1949 une très légère différence d'énergie entre deux niveaux de l'atome d'hydrogène, que le calcul donnait jusque-là pour identiques. Une application plus poussée de la théorie quantique des champs (inventée en 1932 par Pauli et Weisskopf) permit de rendre compte de cette différence — au prix d'une nouvelle difficulté conceptuelle sur laquelle nous ne pouvons insister, et qui laisse aux physiciens un sentiment d'insatisfaction.

Enfin et surtout, il reste encore le problème fondamental des particules élémentaires — électron, neutrino, méson μ, méson π méson K, nucléons, hypérons, dont la liste, déjà longue, s'allonge toujours. Pour le moment, ces particules sont étiquetées par leur masse et par un jeu de nombres quantiques : spin, spin isotopique, parité, étrangeté, hypercharge. La situation n'est pas sans rappeler celle où se trouvait Bohr, confronté avec le mystère des termes spectraux des atomes. Y a-t-il, comme le pense entre autres Heisenberg, une équation d'où « sortiraient » ces particules, un peu comme les niveaux d'énergie de l'atome d'hydrogène sortent de l'équation de Schrödinger? Ou bien — version moderne du mythe de Sisyphe, que suggérait récemment un physicien renommé — un « Dieu malin » crée-t-il sans cesse de nouveaux problèmes pour tenir l'homme en haleine!

Ecartons cette hypothèse sans doute fantaisiste; il reste que nul ne saurait dire si la complexité actuelle des choses est réelle, ou seulement apparente.

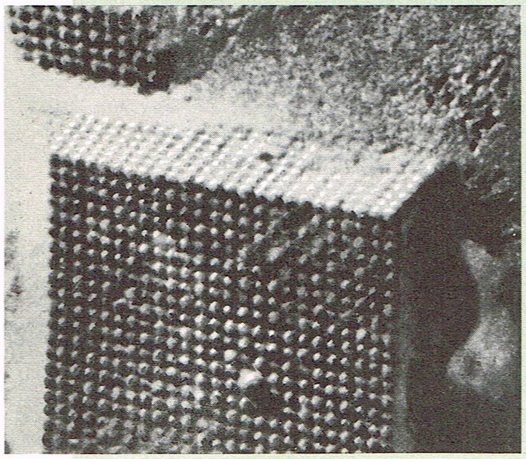
Sans doute trouvera-t-on de nouveaux principes directeurs, et peut-être pourra-t-on, suivant l'antique désir humain, tout ramener à quelques lois? Ces lois seront-elles simples? Peut-être le Larousse de l'an 2000 donnera-t-il une réponse à ces questions.

---

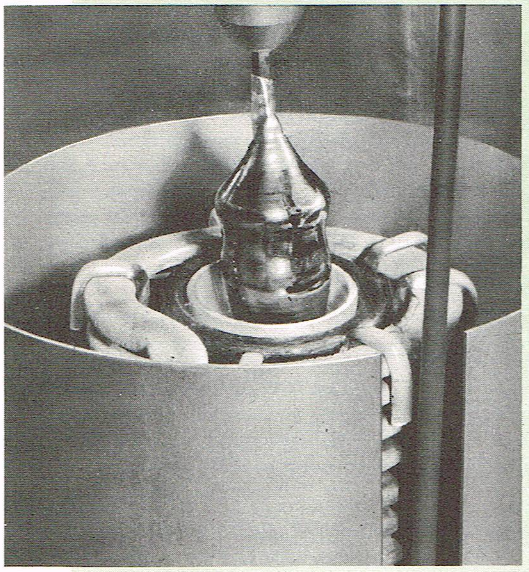
(1) Traduction dans le langage quantique de la propriété qu'ont un système et son image dans un miroir d'obéir aux mêmes lois physiques. Propriété qui paraît presque évidente (« Je ne peux croire que Dieu soit gaucher ») et qui pourtant, dans certains cas, est fausse.



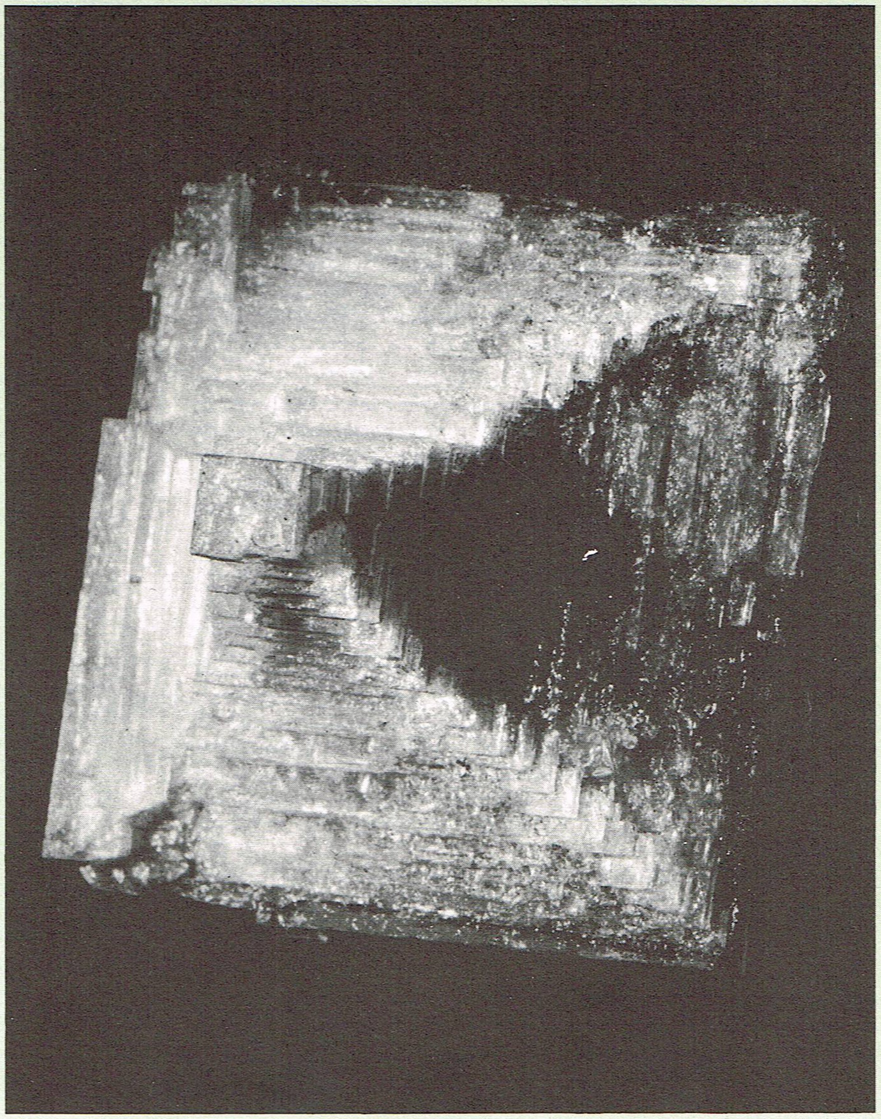
Cristaux de chalcopyrite.



Virus de la mosaïque du tabac (on distingue la triple périodicité du cristal).



Cristal de germanium en cours d'affinage au creuset.



Cristal de chlorure de sodium, ou sel commun (v. p. 312).

Phot. Bariaud, C. S. F., C. N. R. S.

L'ARCHITECTURE DE LA MATIÈRE

par **Maurice BERNARD**

Dès l'âge le plus tendre, nos sens nous apprennent que le monde extérieur est formé de solides et de fluides ; la notion d'objet solide, issue de la sensation du toucher, est associée à l'idée de cohésion, de dureté. Il y a plusieurs millions d'années, l'espèce humaine, par un lent processus, a émergé progressivement des lignées animales. Il est naturel de penser que cette ascension est liée à une conscience de plus en plus aiguë de la notion d'objet ; la pierre que brandit le primate, sans doute sa première arme, préfigure la future domination de la nature par l'homme.

Parce que certains objets sont solides, ils deviennent massues, outils, demeures. Aujourd'hui, l'homme moderne dispose d'une gamme immense d'objets, qui forment la base matérielle de notre civilisation. L'homme contemporain extrait les matériaux que lui offre la nature ou bien les fabrique à partir de matières premières ; il les transforme, les façonne, leur donne des propriétés déterminées à l'avance et les combine judicieusement pour produire les milliers d'objets qui nous entourent : le paquebot et la machine à laver, le transistor et le satellite.

Comment l'homme a-t-il peu à peu pénétré le domaine des solides ? Depuis l'Antiquité jusqu'à la Renaissance, les propriétés les plus évidentes des solides (par exemple, la dureté des métaux, l'éclat du diamant, la transparence de certains minéraux ou leur coloration, etc.) sont simplement admises. Elles sont « expliquées » suivant la démarche propre à la scolastique : de même que l'opium possède la vertu dormitive, de même les métaux, par exemple, ont le pouvoir de la dureté ou celui de la réflexion métallique. La seule arme de la pensée scientifique demeure la classification : la connaissance des solides, comme toute connaissance scientifique, reste descriptive et empirique. Cela n'empêche pourtant pas l'ingéniosité et la patience humaines de bâtir des cathédrales ou de peindre des retables ; c'est que l'empirisme suffit pour estimer l'effort maximal que peut supporter un pilier, ou pour mélanger les pigments qui produiront une couleur.

*Bientôt la mécanique newtonienne et le calcul différentiel ouvrent la porte à la connaissance scientifique. Pourtant la théorie des solides reste dans l'enfance jusqu'à la fin du XIX*e *siècle.*

Les raisons d'une si tardive éclosion sont simples. Dès la fin du XVIIIe siècle, les propriétés de symétrie de certains solides, les cristaux, sont reconnues et étudiées ; mais comment pourraient-elles être comprises, ou simplement interprétées, tant que l'hypothèse atomique ne permet pas de reconnaître en tout solide un édifice construit avec des noyaux et des électrons ? Les progrès de l'optique conduisent à distinguer les solides transparents des solides opaques ; l'indice optique des premiers ou le pouvoir absorbant des seconds peuvent être mesurés ; mais comment comprendre l'un ou l'autre tant que l'on ignore par quel mécanisme la lumière traverse sans affaiblissement un cristal transparent, tant que l'on ne sait pas ce que devient la lumière dans un solide absorbant ? Les lois de l'électricité conduisent à diviser les solides en deux classes distinctes, isolants et conducteurs : une classification et rien de plus, tant que la physique quantique ne viendra pas nous en donner une explication somme toute assez simple.

De la Renaissance au XIXe siècle la connaissance descriptive et phénoménologique des solides suit les progrès de l'électricité, de la chaleur, de l'optique, etc. Mais la véritable physique des solides, en tant que discipline jouissant d'autonomie et fondée sur les principes de base, est très récente ; il ne pouvait pas en être autrement.

Nous nous attacherons d'abord à définir ce qu'on appelle les « solides ». Nous verrons que certains d'entre eux, les cristaux, jouent un rôle primordial, du point de vue théorique comme du point de vue pratique ; cela résulte des remarquables propriétés géométriques qu'ils présentent.

L'objet étudié, un solide formé d'un ou de plusieurs cristaux, étant défini, une question se pose : **est-il possible de fonder sur quelques principes simples une théorie qui puisse rendre compte de toutes les propriétés des solides réels ?** *La théorie moderne des solides bâtie à l'aide de la mécanique quantique a cette ambition. Pourtant, les difficultés mathématiques rencontrées sont telles, que, chemin faisant, de nombreuses approximations seront nécessaires. Ainsi le mouvement des atomes d'un solide et le mouvement de ses électrons seront considérés séparément. Le mouvement des atomes permet d'expliquer la chaleur spécifique des solides, leur dilatation thermique, leur conductivité thermique, etc. Le mouvement des électrons conduit à la théorie dite de l' « approximation des bandes », dont les applications sont spectaculaires et qui est devenue depuis trente ans un outil remarquable pour l'explication d'une foule de propriétés des solides : conductivité électrique, propriétés optiques, magnétiques, etc. Par exemple, nous verrons que les classifications établies empiriquement depuis longtemps s'expliquent naturellement. Nous terminerons en montrant comment ces connaissances ont engendré une foule d'applications qui affectent tous les jours davantage notre genre de vie.*

solides et cristaux

Tous les solides ont en commun cette propriété de dureté et de cohésion qui les différencie des liquides et des gaz. Mais cette définition vague et qualitative n'est pas satisfaisante. Nous savons aussi, depuis les succès de l'hypothèse atomique, que toute matière est formée de particules, les atomes, et il est tout naturel de chercher une définition fondée sur cette réalité microscopique.

l'état solide

Dans un gaz, les atomes, éventuellement groupés en molécules caractéristiques du gaz considéré, sont à des distances les uns des autres considérables par rapport à leurs propres dimensions; ils sont animés de mouvements et de chocs incessants; ainsi s'explique la faible densité d'un gaz, de même que son aptitude à remplir tout récipient que l'on met à sa disposition et à exercer sur ses parois une pression d'autant plus grande qu'on le comprime davantage. **Dans un liquide,** les atomes ou les molécules se « touchent » et glissent les uns contre les autres sans se pénétrer ni beaucoup se séparer; leurs attractions mutuelles sont suffisantes pour qu'atomes ou molécules restent en contact, mais sont trop faibles pour que leurs positions respectives restent fixes au cours du temps; les liquides sont donc à peu près incompressibles sans pourtant posséder de forme propre. Au contraire, **dans un solide,** les interactions sont assez grandes pour que les atomes occupent dans l'espace des positions fixes les uns par rapport aux autres; les atomes sont tous liés de proche en proche d'une manière permanente. Il en résulte d'ailleurs que la notion de molécule disparaît : un solide apparaît comme un assemblage rigide d'atomes formant une sorte de molécule géante; d'où la rigidité des solides et l'existence d'une forme extérieure définie. En fait, cette fixité n'est pas absolue : chaque atome, comme nous le prouverons plus loin, oscille autour d'une position d'équilibre sans beaucoup s'en éloigner, et cette position moyenne reste fixe dans le temps. Si l'on élève la température d'un solide, les amplitudes des oscillations atomiques croissent jusqu'au moment où, à une température bien définie, elles sont si grandes que toute rigidité disparaît : le solide se transforme en un liquide — c'est la fusion — ou en un gaz — c'est la vaporisation. En résumé, par rapport à un système de coordonnées lié à un solide, chaque

atome de ce solide a une position moyenne définie par trois coordonnées fixes.

un état dans l'état : le milieu cristallin

Les propriétés « cristallines » de certains solides sont connues depuis très longtemps. Par exemple, de nombreux minéraux présentent des facettes naturelles ou des faces de clivage parallèles à certaines directions de plans; ces directions de plans se déduisent les unes des autres par des opérations de symétrie. Les **cristaux** connus ont depuis longtemps été classés suivant la nature des symétries que présentent non seulement leurs formes extérieures, mais plus généralement l'ensemble de leurs propriétés. La théorie mathématique des groupes a permis, il y a une vingtaine d'années, de rendre complète et rigoureuse cette classification, qui avait été entreprise bien avant que l'hypothèse atomique permette de soupçonner tout solide d'être formé d'atomes. Pourtant, seule l'architecture atomique permet de comprendre l'état cristallin, et en particulier l'existence de lois de symétrie. Supposons qu'un solide, c'est-à-dire une collection d'atomes rigidement liés les uns aux autres (telle que celle que nous avons définie plus haut), soit homogène, de sorte qu'une quelconque petite région du cristal ait, par définition, les mêmes propriétés physiques que toute autre région de mêmes dimensions (ce solide est assez grand pour que les effets de surface y soient négligeables). Ses propriétés physiques dépendent de la manière dont les atomes de chaque région sont arrangés; ces propriétés ne seront identiques que si, dans chaque région, l'arrangement des atomes est le même; on a donc, dans ce solide, un motif répété identiquement et indéfiniment dans trois directions non coplanaires. **On dit qu'un tel solide est un cristal,** et nous verrons plus tard comment les lois de symétrie résultent de cette conception.

tout solide est-il donc un cristal?

Imaginons qu'à un instant donné les atomes qui forment un liquide soient figés aux places qu'ils occupent à cet instant; ces positions sont aléatoires et ne présentent aucune régularité, aucune périodicité; nous avons fabriqué une structure rigide dans laquelle il n'y a pas de motif régulièrement répété, c'est-à-dire un « solide » qui n'est pas un cristal. On dit qu'un tel système est dans l'**état vitreux** (état liquide de viscosité extrêmement élevée); de tels systèmes existent, les verres en sont des exemples. Un solide dans l'état vitreux est-il homogène? Si une région comprend un nombre suffisant d'atomes répartis statistiquement, c'est-à-dire si le désordre est assez grand, on aura une quasi-homogénéité, au moins à une échelle grande par rapport à la distance moyenne entre deux atomes. A l'heure actuelle, on considère que l'état vitreux est plus proche de l'état liquide que de l'état solide. Bien que ce point puisse être discuté, nous n'étudierons pas les verres.

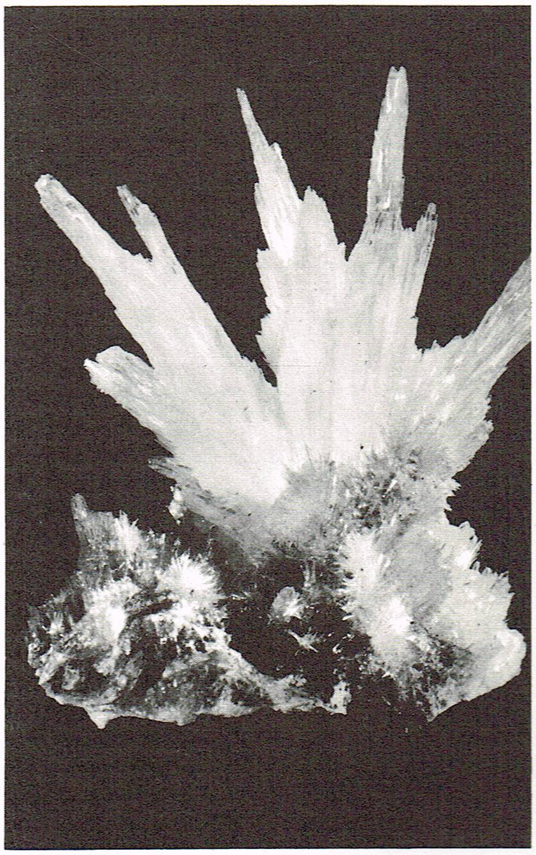
Peut-on en conclure que tous les solides que nous considérerons seront des cristaux? La nature ne se laisse pas enfermer aisément dans des définitions simples. Dans certaines structures fibreuses, comme celle de la cellulose, un même motif est répété indéfiniment dans une seule direction, formant des chaînes linéaires plus ou moins parallèles, plus ou moins rigidement liées les unes aux autres : en quelque sorte, un état cristallin dans une direction et un état vitreux dans un plan perpendiculaire. Les **molécules à longues chaînes** (que nous avons vues dans le premier tome) jouent un rôle important dans les tissus vivants ou dans les produits de synthèse de la chimie organique (plastiques, caoutchouc, etc.). Nous n'en parlerons pas ici.

Une autre objection à propos du caractère général de l'état cristallin vient encore à l'esprit : **il ne semble pas, à première vue, qu'un échantillon de métal (un « vrai solide » comme du fer) présente les caractéristiques extérieures d'un cristal!** Et pourtant, si on le chauffe, aucun ramollissement progressif n'est observé, sa température de fusion est bien définie, et il faut, à cette température, pour le faire passer à l'état liquide, lui fournir une certaine quantité de chaleur; tous ces caractères le différencient nettement d'un verre. Un examen au microscope permet de lever l'énigme : on observe un enchevêtrement de tout petits cristaux de l'ordre de quelques centièmes ou du millième de millimètre de long; on sait d'ailleurs aujourd'hui obtenir des cristaux uniques, ou **monocristaux,** d'un très grand nombre de métaux et métalloïdes.

On voit que la notion de cristal, plus générale qu'il ne paraît, est à la



Le verre, même s'il s'agit du « cristal » de Bohême, est un faux cristal... mais cette masse informe de calcite est une matière cristalline.



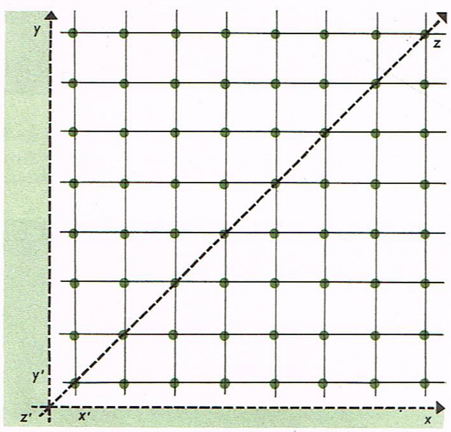
Phot. A. Papillon et P. Memin, C. N. R. S.

base même de l'état solide. Avant de voir comment on peut échafauder la théorie quantique des cristaux, expliquer leur cohésion et bien d'autres propriétés, nous allons d'abord déduire de la définition des cristaux quelques conséquences importantes.

le cristal parfait

Par définition, un cristal parfait est un ensemble d'atomes disposés régulièrement dans l'espace de façon triplement périodique. Les cristaux réels, comme nous le verrons plus loin, diffèrent plus ou moins des cristaux parfaits; nous négligerons pour le moment cette différence. Sachant comment sont disposés les atomes d'un cristal, nous devons être capables de dire d'avance les symétries de ce cristal; c'est là pure question de géométrie.

Assimilons les atomes à de toutes petites billes et disposons-les d'abord, comme



Un réseau plan carré et ses éléments de symétrie.

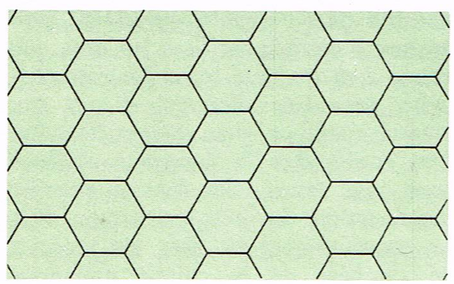
Les axes $x'x$ et $y'y$ parallèles au côté du carré et l'axe $z'z$ parallèle à sa diagonale sont des axes de rotation d'ordre 2 (rotation de $360°/2 = 180°$); l'axe perpendiculaire à la figure (non représenté) est un axe de rotation d'ordre 4 (rotation de $360°/4 = 90°$).

il est indiqué sur la figure ci-dessus, aux sommets d'un réseau plan carré indéfini. Nous avons construit un cristal plan à deux dimensions.

Quels sont ses « éléments de symétrie »? On peut, par exemple, le faire tourner d'un quart de tour autour d'un axe perpendiculaire au plan de la figure : on obtient un arrangement identique à l'arrangement initial. On dit que le cristal considéré possède un axe de rotation d'ordre 4 (c'est-à-dire rotation d'angle $360°/4$); un tel axe est appelé un « élément de symétrie ». De même, une rotation d'un demi-tour autour de l'un des trois axes $x'x$, $y'y$, $z'z$ respectivement parallèles aux côtés du carré et à sa diagonale laisse le cristal inchangé; on dit que les axes $x'x$, $y'y$, $z'z$ sont des axes de rotation d'ordre 2 (rotation d'angle $360°/2$).

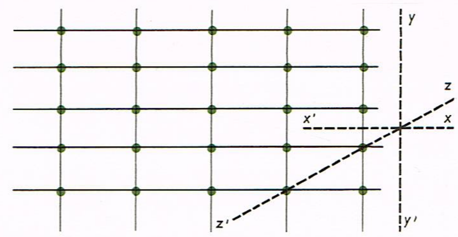
Le cristal plan obtenu en disposant les atomes aux sommets d'un réseau rectangulaire n'a pas les mêmes éléments de symétrie que le cristal de la figure précédente; par exemple, l'axe perpendiculaire au plan de la figure n'est plus un axe de rotation d'ordre 4, mais seulement d'ordre 2; les axes $x'x$ et $y'y$ sont toujours des axes de rotation d'ordre 2, tandis que l'axe $z'z$ n'a plus rien de remarquable.

Il existe aussi des axes de rotation d'ordre 3 ou d'ordre 6, comme on peut le constater immédiatement en examinant



Le réseau plan hexagonal : l'axe perpendiculaire au plan de la figure (non représenté) est un axe d'ordre 6 (rotation $360°/6 = 60°$).

le réseau hexagonal de la figure ci-dessus. On peut démontrer qu'un axe de rotation d'ordre 5 (rotation de $360°/5 = 72°$) ou d'ordre 7 (rotation de $360°/7$) n'est



Un réseau plan rectangulaire et ses éléments de symétrie.

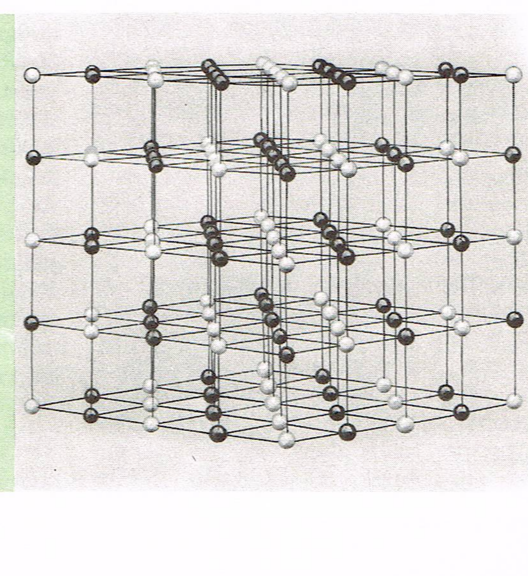
Les axes $x'x$ et $y'y$ parallèles aux côtés de rectangle sont toujours des axes de rotation d'ordre 2, tandis que l'axe $z'z$ parallèle à la diagonale n'est plus élément de symétrie, l'axe perpendiculaire au plan de la figure est encore un axe de rotation d'ordre 2 (au lieu de 4 pour le réseau carré).

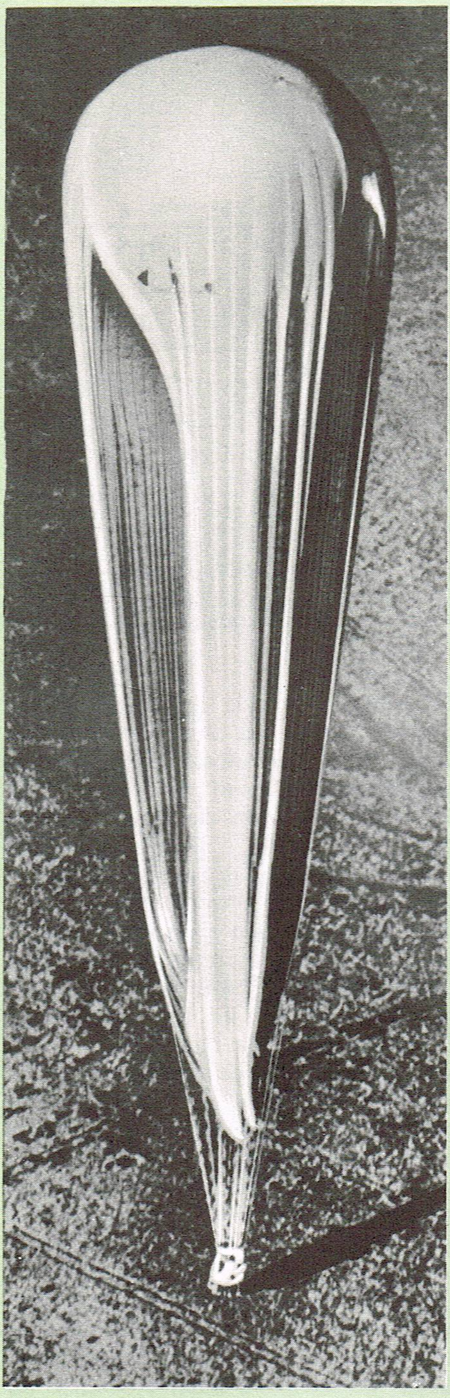
pas compatible avec un arrangement périodique plan indéfini; plus précisément, on peut montrer que les seuls axes de rotation possibles sont d'ordre 2, 3, 4, 6. On définit aussi comme éléments de symétrie la symétrie par rapport à un point et la symétrie par rapport à un plan.

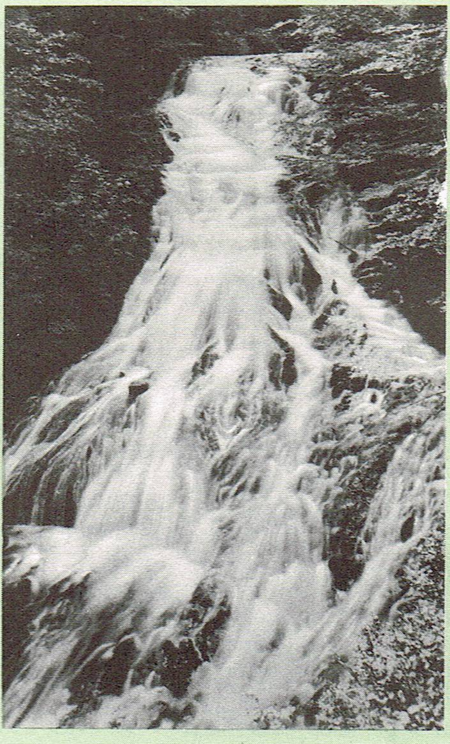
Pour la simplicité de l'exposé, nous avons jusqu'à présent considéré des cristaux à deux dimensions. Dans l'espace à trois dimensions, disposons une série de réseaux carrés, les uns au-dessus des autres, de sorte que les atomes soient disposés suivant un réseau cubique à trois dimensions.

On peut encore définir des éléments de symétrie; par exemple, les trois axes parallèles aux arêtes du cube sont des axes de rotation d'ordre 4, alors que la direction parallèle à la diagonale du cube est un axe de rotation d'ordre 3. Si nous avons assez de patience et d'imagination, nous pourrons déterminer les différentes façons de disposer dans l'espace des billes suivant un arrangement

Le réseau cubique : sel commun avec ses atomes de chlore et de sodium (billes respectivement noires et blanches).







État liquide :
les molécules glissent
l'une sur l'autre.

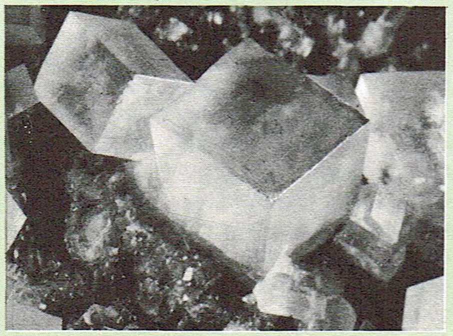


État vitreux :
état liquide
de viscosité extrême.

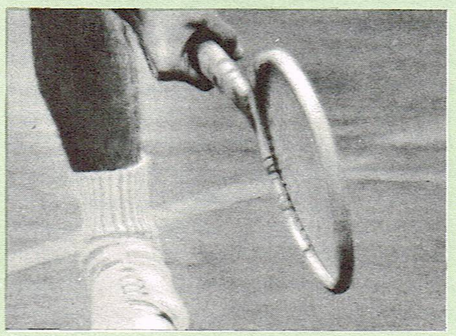
LES PRINCIPAUX ÉTATS DE LA MATIÈRE

État gazeux :
les molécules n'interagissent
guère, tant leurs distances
mutuelles sont grandes.

État cristallin
(motif répété indéfiniment
dans tout l'espace).



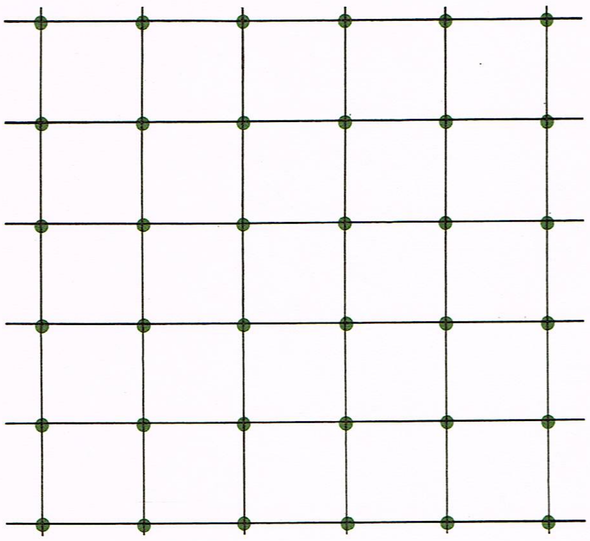
Chaînes fileuses
(la répétition du motif n'a
lieu que le long d'une direction).



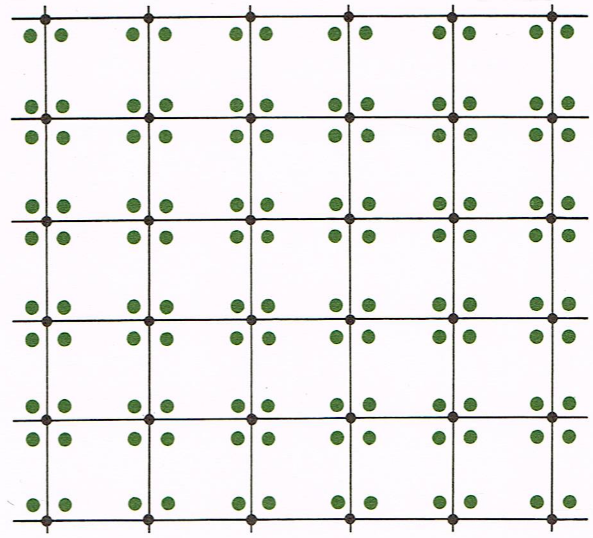
Microcristaux
(enchevêtrement de cristaux
de quelque 10^{-9} centimètre).

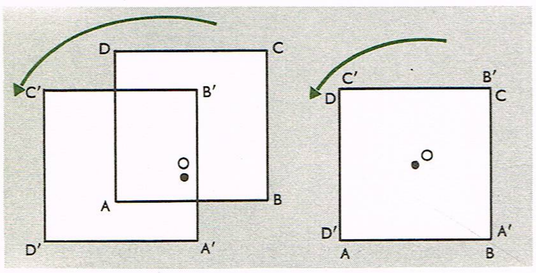


Phot. Neurdein, Giraudon, Bariaud, Presse Sports.



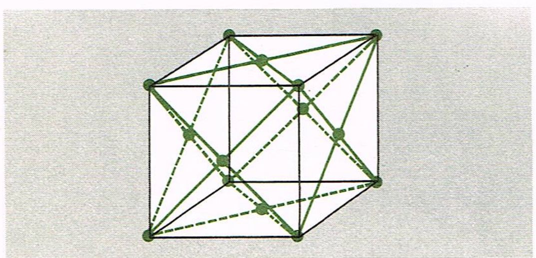
Deux exemples de réseau carré simple





Position des éléments de symétrie.

Dans la partie gauche de la figure, une rotation de 90° autour d'un axe perpendiculaire au plan de la figure et la perçant en un point O quelconque transforme le carré ABCD en un carré A′B′C′D′; ces deux carrés ne coïncident pas, mais peuvent être superposés par une translation parallèle au vecteur $\overrightarrow{A'B}$. Dans la partie droite de la figure, l'axe de rotation d'ordre 4 passe par le centre O du carré : A′B′C′D′ coïncident exactement avec BCDA.



Les atomes d'un cristal de cuivre sont répartis aux sommets d'un réseau cubique à face centrée; un calcul simple montre que si a est le côté du cube, il y a $4/a^3$ atomes de cuivre par unité de volume. La théorie atomique montre aussi qu'un atome-gramme du cuivre, c'est-à-dire un échantillon de 63 grammes, contient N atomes (N, appelé nombre d'Avogadro, vaut 6 10^{23}); comme la masse spécifique du cuivre est de 8,9 g/cm, il y a $\frac{8,9 \times 6.10^{23}}{63}$ atomes par unité de volume; et égalant ce nombre à $4/a^3$ on en déduit : $a \simeq 3,4\,10^{-8}$ cm = 3,4 angströms (l'angström est la cent-millionième partie d'un centimètre; les dimensions des atomes sont voisines de l'angström).

régulier triplement périodique; mais les arrangements possibles sont innombrables; peut-on trouver une classification qui simplifie la description des arrangements possibles?

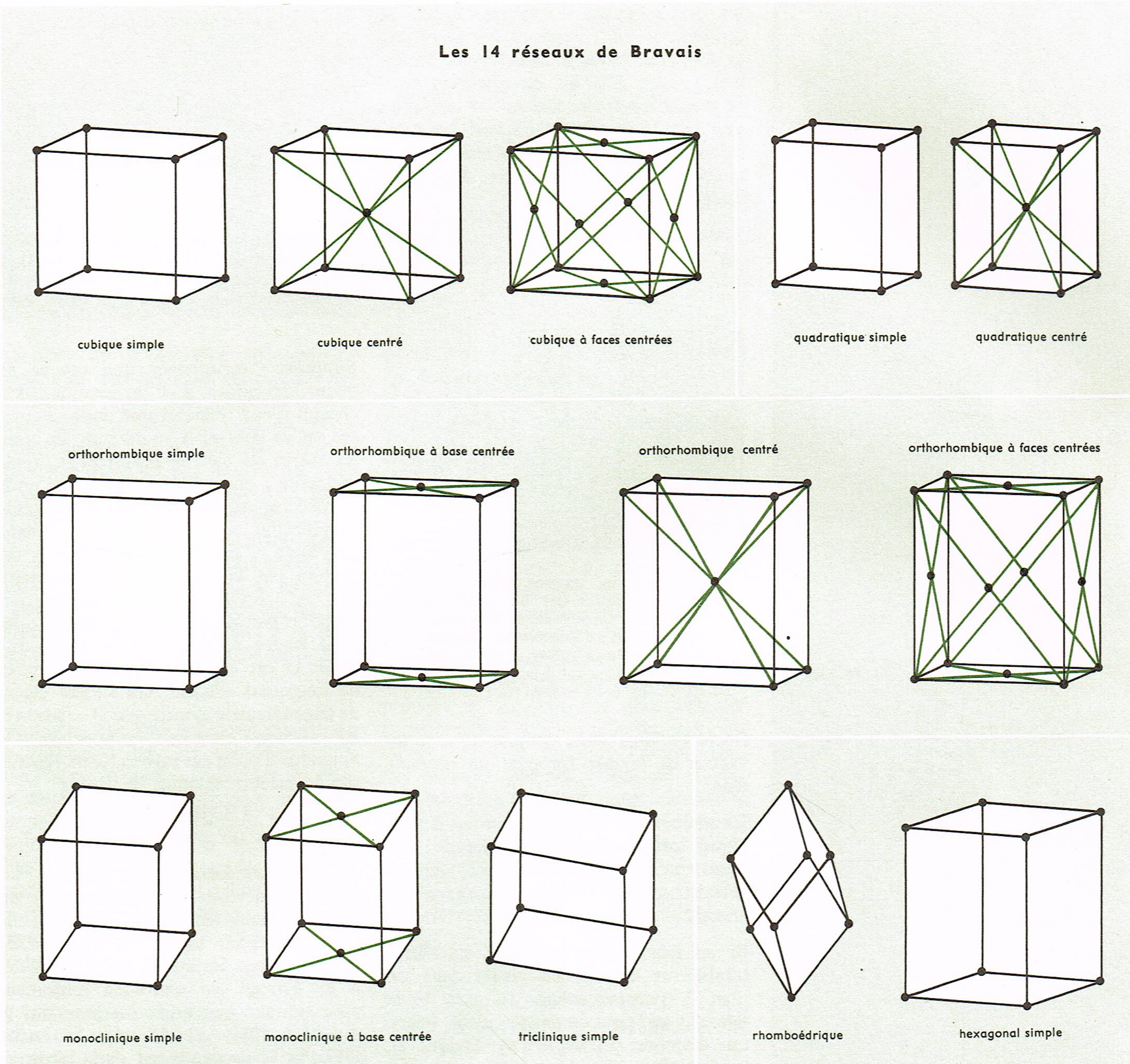
Prenons pour exemple les deux arrangements plans de la figure ci-dessus; celui de droite se déduit de celui de gauche en remplaçant une bille sphérique par un ensemble de quatre billes. Mais si nous faisons abstraction de la nature de ce motif élémentaire, nous pouvons considérer les deux arrangements comme identiques : on dira que l'un et l'autre appartiennent au réseau carré simple. En somme, nous pouvons classer les arrangements d'atomes d'après la nature du réseau selon lequel un motif donné, quel qu'il soit, se répète indéfiniment dans l'espace. Cette répétition dépend de trois translations élémentaires, c'est-à-dire de trois vecteurs; le réseau d'un arrangement d'atomes ne dépend que des trois vecteurs « bases » de ces translations. On démontre que tout réseau appartient nécessairement à l'un des quatorze réseaux de Bravais représentés page ci-contre.

Les trois réseaux cubiques, cubique simple, cubique centré et cubique à faces centrées, ont tous trois les mêmes symétries, à savoir celles du cube; on dira que ces trois réseaux appartiennent au système cubique; pourtant, la répétition d'un même motif ne sera pas la même dans l'un ou l'autre de ces trois réseaux, puisqu'ils diffèrent par leurs vecteurs de base; on voit apparaître la nécessité de deux classifications : celle des réseaux cristallins ou réseaux de Bravais, au nombre de quatorze, et celle des systèmes cristallins, au nombre de sept.

Nous arrivons à la conclusion qu'**un cristal parfait est construit à partir d'un motif atomique élémentaire répété indéfiniment par les translations d'un réseau de Bravais.** Jusqu'à présent, nous n'avons pas cherché à savoir quel était le motif atomique et nous avons assimilé ce motif à une bille. En pratique, il peut exister des cristaux dont le motif soit formé d'un élément très voisin d'une bille, mais ce cas est plutôt exceptionnel. Si, maintenant, nous tenons compte de la nature de ce motif, de ses éléments de symétrie, en les combinant avec les éléments de symétrie du réseau, nous arrivons à une classification des cristaux en trente-deux classes, qui diffèrent les unes des autres par la nature des éléments de symétrie caractérisant chaque classe.

Tous les éléments de symétrie que nous avons considérés jusqu'à présent ne sont définis qu'en direction, et non en position; par exemple, lorsque nous avons remarqué que le cristal de la figure (p. 312) admet un axe perpendiculaire au plan de la figure comme axe de rotation d'ordre 4, nous n'avons pas spécifié en quel point cet axe perçait le plan de la figure; en d'autres termes, deux positions d'un cristal déduites l'une de l'autre par une opération de symétrie sont considérées comme identiques si elles peuvent être superposées par une translation. Si, dorénavant, nous tenons compte non seulement de la nature des éléments de symétrie, mais encore de leur position exacte dans le cristal, nous aboutissons à une nouvelle classification;

Les 14 réseaux de Bravais



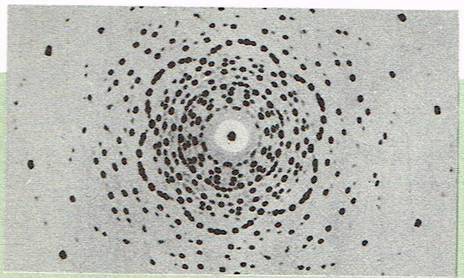
on démontre qu'il existe 230 groupes, dits « groupes d'espace », qui diffèrent les uns des autres par **la nature et la position** des éléments de symétrie. L'examen détaillé des 7 systèmes cristallins, des 14 réseaux de Bravais, des 32 classes de symétrie et des 230 groupes d'espace, l'étude de leurs rapports mutuels dépassent largement le cadre de cet ouvrage; **la théorie des groupes permet heureusement de rendre logiques et rigoureuses toutes les classifications de la cristallographie.**

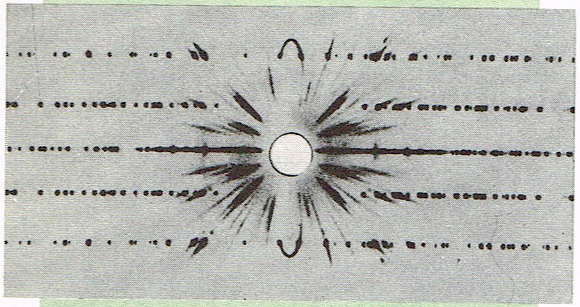
Pour nous, le résultat essentiel est qu'un cristal parfait est entièrement déterminé par la connaissance d'un motif élémentaire, c'est-à-dire par la donnée d'un assemblage défini et généralement simple d'atomes, ce motif étant répété indéfiniment dans les trois directions d'espace à partir des trois translations définissant le réseau. Réciproquement, tout cristal parfait que l'on peut imaginer dans l'espace à trois dimensions (l'espace « réel » où nous vivons) trouvera nécessairement sa place dans le système de classification cristallographique que nous avons brièvement esquissé plus haut.

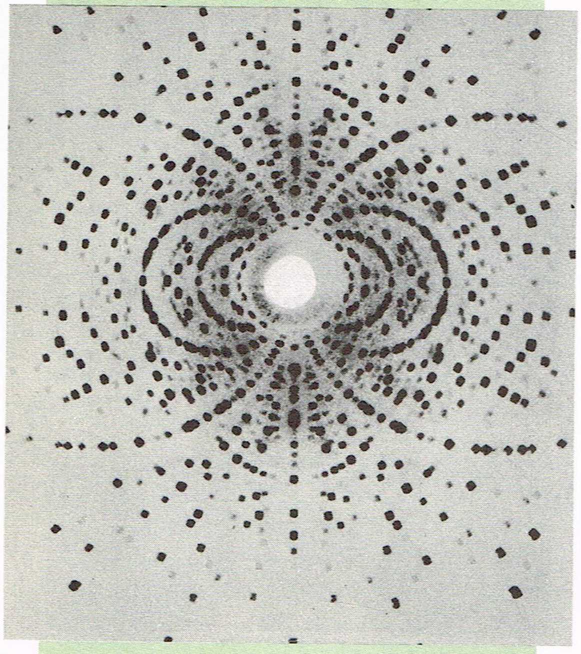
A ce moment, le lecteur curieux a sur les lèvres une question : **Peut-on, en présence d'un cristal, savoir le type auquel il appartient?** Depuis le début du siècle, c'est-à-dire depuis la découverte de la diffraction des rayons X par les cristaux, on peut répondre affirmativement à cette question.

les rayons X révèlent la structure des cristaux

On sait depuis la fin du XIX[e] siècle que les rayons X (rayonnements émis par une plaque métallique bombardée par un faisceau d'électrons rapides) sont des ondes électromagnétiques, tout comme







LES RAYONS X RÉVÈLENT LA STRUCTURE DES CRISTAUX

De haut en bas :

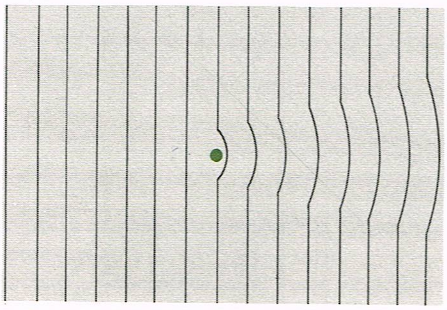
cliché de Laue du béryl montrant sa symétrie sénaire;

diagramme de cristal tournant;

diagramme de Laue de la topaze (symétrie orthorhombique).

les ondes de radio ou les ondes de lumière; à la différence de ces dernières, la longueur d'onde des rayons X est très courte. Elle est de quelques angströms, c'est-à-dire du même ordre de grandeur que les distances interatomiques, et cette coïncidence remarquable a, comme nous allons le voir, des conséquences capitales.

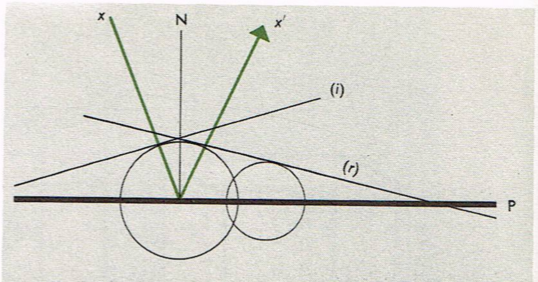
Lorsqu'un faisceau de rayons X tombe sur un atome, le rayonnement est diffracté en tous sens, comme le sont les vagues de la mer frappant un obstacle fixe.



Lorsque les vagues d'un plan d'eau rencontrent un obstacle fixe, leur mouvement régulier est modifié : de nouvelles vagues de directions différentes apparaissent. On dit que les ondes se sont diffractées.

Le phénomène de diffraction est-il analogue lorsque le faisceau tombe non plus sur un atome isolé, mais sur un grand nombre d'atomes formant un cristal?

Si un faisceau de rayons X parallèles tombe sur des atomes situés dans un plan P (perpendiculaire au plan de la figure), on peut montrer qu'il existe une direction privilégiée, symétrique de la direction d'incidence par rapport à la normale au plan réticulaire P, qui correspond à une onde réfléchie intense, puisque les ondes diffractées par chaque atome dans cette direction sont en phase et se renforcent.



La diffraction par un plan réticulaire : l'onde (de plan d'onde incidente dans la direction *x*) est diffractée par le plan réticulaire P en une onde (de plan d'ordre *r*) de direction *x'*.

Supposons maintenant une rangée de plans identiques à P et équidistants; chacun d'eux réfléchit une onde. Toutes ces ondes vont-elles se détruire les unes les autres ou bien se renforcer? L'onde incidente i donne naissance à une onde r_1 réfléchie sur le plan P_1 et à une onde r_2 réfléchie sur le plan P_2; le retard de l'onde r_2 par rapport à l'onde r_1 est mesuré par la distance supplémentaire $D = AO + OB$ que la deuxième doit franchir par rapport à la première; les deux ondes se renforceront si la distance D est égale à un nombre entier de longueurs d'onde. Un simple calcul de trigonométrie montre que $D = 2a \sin \theta$. La condition cherchée, en appelant λ la longueur d'onde des rayons X, est fournie par la célèbre relation de Bragg

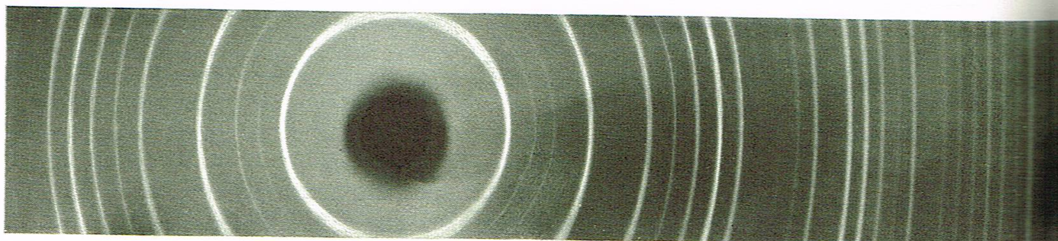
$$2\,a \sin \theta = n\,\lambda,$$

où n est un entier ($n = 1, 2, \ldots$).

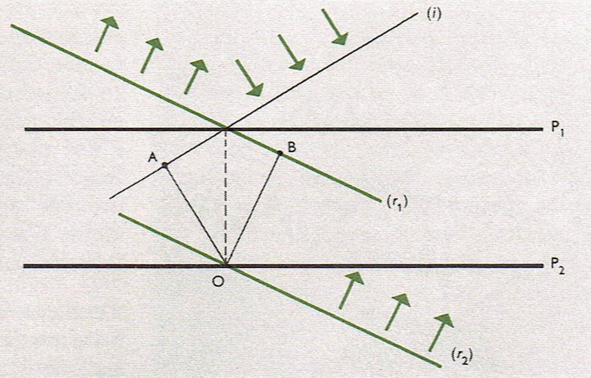
Si cette condition (sans laquelle les ondes se détruisent mutuellement) est satisfaite pour les ondes r_1 et r_2, elle l'est aussi pour toutes les ondes réfléchies par tous les autres plans parallèles à P. Toutes ces ondes se renforcent pour donner une onde intense dans la direction Ox'; on dit que le cristal diffracte le rayonnement dans la direction Ox'.

Dans la formule ci-dessus, n est un nombre entier généralement petit (1 ou 2, par exemple). Comme $\sin \theta$ est compris entre 0 et 1, on voit clairement que la diffraction des rayons X par les cristaux est due au fait que la longueur d'onde λ de ces rayonnements est de l'ordre de

Diagramme Debye-Scherrer du niobate de lithium, composé cubique.



Doc. Durif, Grenier, Martin, du C. N. R. S.



La diffraction de Bragg : l'onde incidente (*i*) est diffractée par le plan réticulaire P_1 en une onde (r_1), puis par le plan réticulaire P_2 en une onde (r_2); la différence de marche entre ces deux ondes est $AO + OB = 2\,a \sin \theta$.

grandeur de la distance séparant les atomes. Si, par exemple, on utilisait des rayons γ ou rayons X de longueur d'onde beaucoup plus courte, la formule précédente montre qu'il faudrait avoir *n* grand (ce qui n'est pas souhaitable pour la qualité des mesures) ou bien avoir un angle θ très petit, c'est-à-dire attaquer le cristal sous une incidence rasante (ce qui se révèle en général assez peu pratique). Si, en revanche, la longueur d'onde λ est plus grande que $2a$, aucune valeur de θ ne peut satisfaire à la relation de Bragg : il n'y a plus de diffraction du tout.

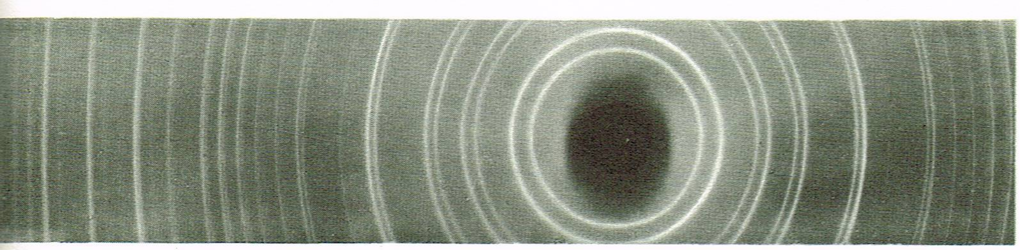
Supposons que l'on fasse une expérience de diffraction avec un rayonnement X de longueur d'onde λ connue et un cristal dont on ignore la distance interatomique *a*; en faisant tourner le cristal, on cherchera la valeur de l'angle θ pour laquelle la relation de Bragg est satisfaite. Cette méthode, dite « méthode du cristal tournant », permet de déduire d'une mesure d'angle la distance interatomique *a*.

Si l'on répète cette expérience avec le même cristal, mais pour diverses directions de plans cristallins, on peut de proche en proche déterminer l'arrangement des atomes dans le réseau du cristal. Cet exemple nous montre, sans entrer dans les détails des techniques radiocristallographiques, comment les rayons X permettent de mesurer avec précision l'architecture des atomes dans les cristaux. A la diffraction des rayons X se sont ajoutées plus récemment celle des électrons et celle des neutrons; grâce à l'arsenal complet de ces techniques, nous pouvons admettre que tout cristal est géométriquement connu.

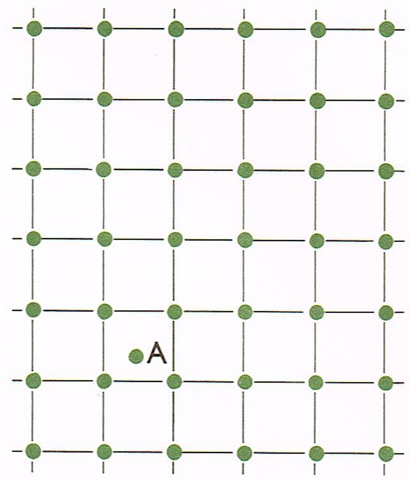
Ce résultat est d'une importance capitale. Voici, par exemple, un morceau de cuivre; nous voulons prédire ses propriétés physiques. La cristallographie nous donne avec une grande précision la position de tous les atomes de cuivre formant ce « cristal »; comme nous le verrons plus loin, nous pouvons nous appuyer sur ces données expérimentales précises et indiscutables pour calculer théoriquement telle ou telle propriété du cuivre. Si l'on ignorait l'architecture des atomes dans un cristal de cuivre, on devrait, adoptant une démarche inverse, chercher le plus stable parmi tous les arrangements possibles. Cette façon de faire se heurterait à d'immenses difficultés de calcul; la cristallographie, heureusement, nous permet d'en faire l'économie.

cristaux parfaits et cristaux réels

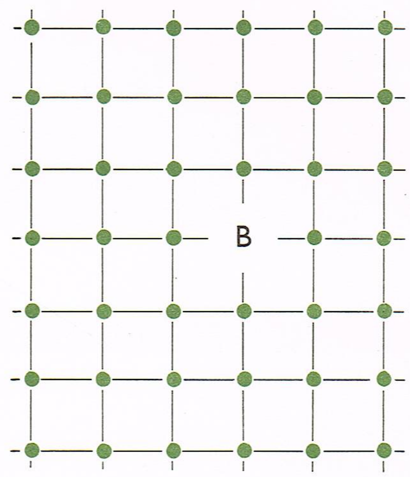
Les cristaux réels — ceux que la nature a fabriqués ou ceux que nous synthétisons en laboratoire — ne sont pas les arrangements géométriques parfaits dont nous avons parlé; ils en diffèrent par ce que l'on appelle les **défauts cristallins.** Par exemple, un cristal réel est fini, limité par des surfaces où l'arrangement



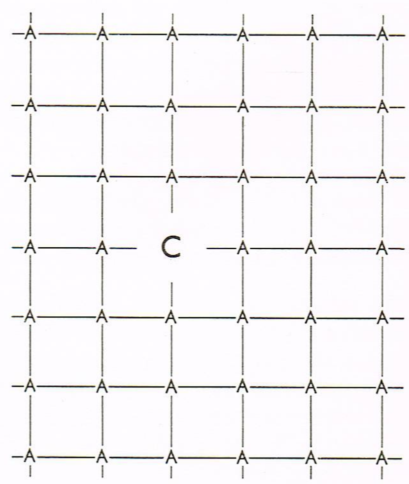
Quelques défauts ponctuels



l'atome A est un défaut « interstitiel »

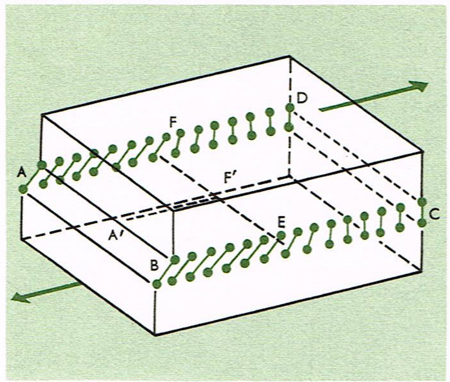


la place B vacante est une « vacance » ou lacune



l'atome C est une impureté « substitutionnelle »

Un défaut de réseau linéaire est appelé une dislocation : ABCD est le plan de glissement, la partie ABEF a glissé dans la direction de glissement A′F′ par rapport à la partie FEDC, la ligne EF est une ligne de dislocation.



Les dislocations jouent un rôle très important dans l'explication de la croissance des cristaux : on voit ici une image, obtenue par microscopie électronique, montrant la croissance en spirale d'un cristal de paraffine à partir d'une dislocation du cristal sous-jacent.



périodique des atomes est rompu ; les surfaces *sont* des défauts inévitables ; ou encore, dans un cristal, un atome manque en un nœud du réseau : c'est une **lacune ;** au contraire, si un atome est en excès à une place normalement vide, c'est un **interstitiel ;** à la place normalement occupée par un atome de cuivre, un atome d'argent est une **impureté substitutionnelle** dans le réseau du cuivre.

Tous ces défauts conservent à un cristal ses caractères essentiels : on dit que l'on a affaire à un « cristal unique » ou « monocristal ». Si nous préparons, par exemple, un échantillon métallique par une fusion normale, sans précautions particulières, nous obtenons un solide très compliqué formé de petits cristaux enchevêtrés ; chaque petit cristal est un cristal unique avec ses propres défauts de réseau ; les surfaces qui limitent deux microcristaux adjacents sont appelées « joints ».

La démarche générale de la physique des solides consiste à chercher d'abord les propriétés des cristaux parfaits, puis les propriétés des défauts de réseau, et de considérer les propriétés réelles des cristaux réels comme la superposition des propriétés des cristaux parfaits et des propriétés des défauts du réseau.

vibrations des atomes dans un cristal

Un cristal réel ne peut être parfait pour la simple raison que ses atomes, à moins qu'il ne soit maintenu à une température très proche du zéro absolu, sont nécessairement animés, comme nous l'avons dit plus haut, de mouvements vibratoires autour de leurs positions moyennes. Seules ces positions moyennes sont exactement situées, en l'absence de tous autres défauts, aux nœuds géométriques du réseau cristallin.

Le moment est venu d'aborder une difficulté : **qu'entend-on exactement par atome, dans un cristal?**

Nous savons, par exemple, qu'un cristal de cuivre est un arrangement périodique formé d'un grand nombre de noyaux de cuivre et d'électrons, à raison de vingt-neuf électrons par noyau de cuivre (puisque le cuivre a pour numéro atomique vingt-neuf).

D'une part, le noyau de cuivre est presque quatre mille fois plus lourd que les vingt-neuf électrons ; d'autre part, nous verrons que la grande majorité de ces vingt-neuf électrons sont, en fait, si rigidement liés au noyau, qu'ils forment avec lui un petit solide à peu près indéformable appelé « ion » ; seuls quelques électrons par atome, les « électrons de valence », ne sont pas rigidement liés au noyau ; nous reviendrons plus loin sur le rôle fondamental que jouent ces électrons. Nous pouvons donc considérer que toute la masse d'un cristal, ou presque, est concentrée dans de petits solides indéformables appelés « ions » ; ces ions exécutent autour de leurs positions d'équilibre des mouvements vibratoires.

Parmi les nombreuses propriétés des solides, un certain nombre d'entre elles dépendent essentiellement des ions et de leurs mouvements et ne dépendent pratiquement pas des électrons. C'est au mouvement des ions et à ces propriétés que ce chapitre est consacré.

mouvement des ions et mode de vibration d'un cristal

Tout au long des deux tomes de cet ouvrage, le lecteur a constaté combien sont importants les phénomènes de résonance. Ainsi une corde tendue, écartée de sa position d'équilibre et abandonnée à elle-même, se met à vibrer à une fréquence bien définie, caractéristique de la tension, de la longueur et de la masse de la corde. Ce système a un seul degré de liberté et ne peut avoir qu'une seule fréquence de résonance ; on l'appelle un « résonateur ». Un système à deux degrés de liberté pourra avoir deux fréquences de résonance différentes, et, de façon générale, nous admettrons que tout système peut avoir autant de fréquences de résonance qu'il possède de degrés de liberté : il

Doc. « Revue de l'aluminium » et Kléber-Colombes.

est équivalent à autant de résonateurs qu'il y a de fréquences de résonance.

En particulier, quels sont les mouvements possibles d'une chaîne linéaire d'atomes régulièrement espacés ? Supposons d'abord tous les atomes rigidement fixés, à l'exception d'un seul, l'atome x, susceptible d'un déplacement unique u_x dans une direction perpendiculaire à la chaîne (*fig.* ci-contre). Ses deux plus proches voisins exerceront sur lui une force de rappel βu_x proportionnelle au déplacement : l'atome va prendre un mouvement sinusoïdal de fréquence f. En réalité, les autres atomes ne sont pas fixes et chacun d'eux est soumis aux forces de rappel que ses voisins exercent sur lui. Si la chaîne linéaire contient $N + 1$ atomes, seuls les atomes extrêmes étant fixes, on a un système à $N - 1$ degrés de liberté, qui doit avoir, *a priori*, $N - 1$ fréquences de résonance, ou, comme on dit, $N - 1$ modes de vibration; ces modes, dans le cas simple que nous considérons, peuvent être calculés assez facilement. La figure ci-dessus montre les positions des atomes de la chaîne à différents instants pour un certain mode caractérisé par une longueur d'onde λ_1 telle que $\lambda_1/2 = L$, où L est la longueur de la chaîne. La figure ci-dessous est relative à un autre mode, caractérisé par une longueur d'onde λ_2 telle que $2\lambda_2/2 = L$. De façon générale, pour le n^ième^ mode, la relation donnant λ_n est $n\lambda_n/2 = L$, où n est un nombre entier ($n = 1, 2, ...$).

Le « vecteur d'onde » $k_n = \frac{2\pi}{\lambda_n}$ (ce vecteur se réduit ici, dans le cas unidimensionnel considéré, à une simple grandeur scalaire) s'écrit

$$k_n = \frac{2\pi}{\lambda_n} = \frac{n\pi}{L}.$$

Il y a $N - 1$ modes distincts; chacun de ces modes a un vecteur d'onde k_n donné par la relation ci-dessus (où n peut prendre les valeurs 1, 2, 3, ... $N - 1$). Le calcul montre que la fréquence de chaque mode est donnée par la formule simple

$$f_n = \frac{1}{2\pi}\left(\frac{4\beta}{M}\right)^{1/2} \sin k_n a,$$

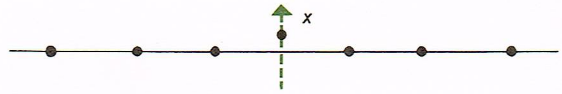
où a est la distance séparant deux atomes consécutifs, M la masse de chaque atome, et β la constante élastique de la force de rappel introduite plus haut.

En fait, à chaque mode correspondent des ondes stationnaires pour les mouvements des atomes; c'est-à-dire que chaque mode peut être considéré comme la superposition de deux ondes progressives, l'une se propageant de gauche à droite et l'autre de droite à gauche, de même fréquence, de même longueur d'onde et de phases telles que les atomes extrêmes restent fixes. Si nous n'avions pas imposé aux atomes extrêmes d'être fixes, nous aurions trouvé pour modes de vibration des ondes progressives ayant des vecteurs d'onde k_n et des fréquences f_n liés encore par la relation précédente.

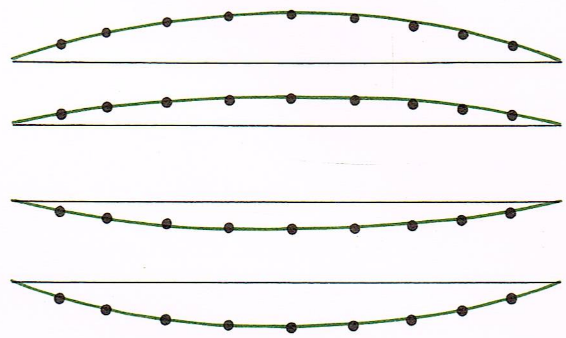
Dans ce qui précède, nous avons envisagé un mouvement transverse perpendiculaire à la chaîne d'atomes; en considérant des mouvements parallèles à la chaîne, on obtient de manière analogue des ondes longitudinales.

Pour un vrai cristal, à trois dimensions, le mouvement des atomes est évidemment beaucoup plus compliqué; les résultats sont pourtant qualitativement comparables à ceux du cas simplifié d'une chaîne linéaire d'atomes. Un cristal contenant N atomes (N étant en pratique très grand) présente divers modes de vibration; pour chacun de ces modes, chaque atome exécute un mouvement sinusoïdal de fréquence f_n, tandis qu'apparaît une oscillation de longueur d'onde λ_n qui correspond à un vecteur d'onde k_n. Chaque atome ayant trois degrés de liberté, le système entier possède 3N degrés de liberté; on montre que le cristal présente 3N modes de vibration, c'est-à-dire contient 3N résonateurs. (En fait, le nombre de modes est un peu inférieur à 3N si l'on suppose les faces du cristal fixes; mais comme N est très grand, nous pouvons négliger cette différence.) Parmi ces modes, N correspondent à des ondes longitudinales et 2N à des ondes transversales.

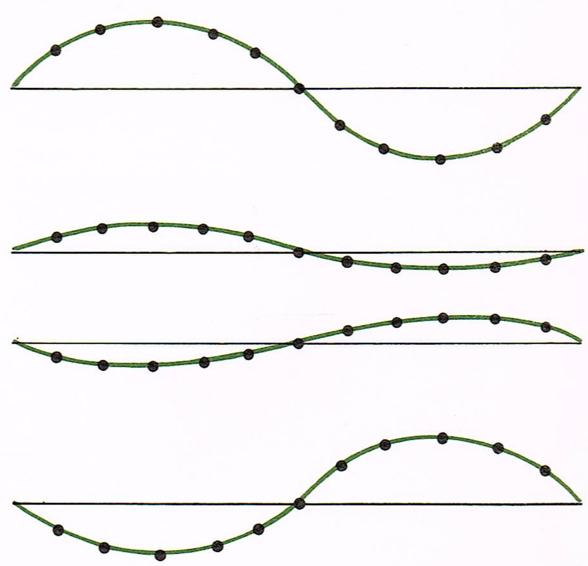
Il serait très satisfaisant que tout mouvement permanent des atomes du cristal correspondît à l'un des 3N modes que nous avons trouvés. Ce n'est malheureusement pas vrai; ou, en tout cas, pas tout à fait vrai. La théorie de l'analyse harmonique fondée sur le très important théorème de la décomposition de Fourier nous permet d'affirmer que tout mouvement permanent des atomes du cristal, s'il n'est pas celui qui correspond à l'un des 3N modes de vibration, est une certaine superposition linéaire des mouvements correspondant à chacun de ces modes (théorème de superposition linéaire). Par conséquent, les modes de vibration d'un cristal permettent de déduire tout mouvement possible des atomes de ce cristal. Un



Sous l'influence de ses voisins supposés fixes, un atome (x) prend un mouvement sinusoïdal de fréquence f, perpendiculaire à la chaîne linéaire d'atomes.



On a représenté, pour un mode donné, les différentes positions occupées à différents instants par les atomes d'une chaîne linéaire dont les extrémités sont fixes. Pour le mode représenté, tous les atomes sont animés de mouvements sinusoïdaux, perpendiculaires à la chaîne, de même fréquence f, et dont les phases respectives sont telles que les atomes, à tout instant, soient situés sur une arche de sinusoïde; la longueur L de la chaîne est donc égale à la moitié de la longueur d'onde λ_1, de cette sinusoïde $\lambda_1/2 = L$.



Dans la figure ci-dessus, on a représenté les positions des atomes d'une chaîne linéaire à différents instants; pour le mode considéré, il y a un nœud supplémentaire au milieu de la chaîne, dont la longueur L est donc égale à deux demi-longueurs d'onde λ_2 : $2\frac{\lambda_2}{2} = L$; on a ensuite $3\frac{\lambda_3}{3} = L$; $4\frac{\lambda_4}{4} = L$, et ainsi de suite.

mouvement non permanent des atomes, par exemple un mouvement amorti, est impossible si le cristal est isolé, ce que nous supposerons pour simplifier.

excitation des modes de vibration d'un cristal et chaleur spécifique des solides

Nous avons accompli un pas important puisque nous connaissons les mouvements possibles. *Mais quel est le mouvement réel?*

On pourrait penser qu'un cristal isolé va rester au repos, ses atomes étant immobiles en leurs positions d'équilibre. Cela n'est vrai qu'au zéro absolu, c'est-à-dire à la plus basse température concevable, à 273 °C au-dessous de la glace fondante. A toute autre température, réelle, les atomes seront animés de mouvements thermiques : les modes de vibration (ou résonateurs) seront excités, et d'autant plus intensément que la température absolue T du cristal sera plus élevée. Trouver le degré d'excitation de chaque résonateur lorsque le cristal est à une température T est un problème difficile auquel seule la mécanique quantique peut apporter une réponse satisfaisante.

où entrent en scène les quanta

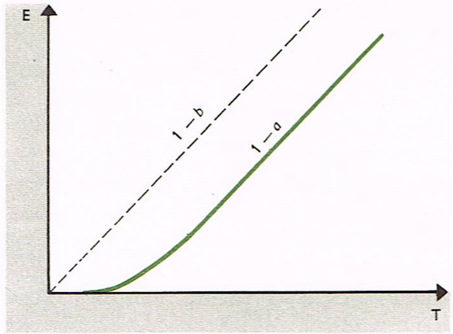
La physique classique conduit à un résultat erroné, en contradiction flagrante avec l'expérience; cette difficulté n'a pas peu contribué, au début de notre siècle, à ruiner les conceptions anciennes et à favoriser l'introduction des quanta. La statistique dite « classique » que Maxwell et Boltzmann avaient élaborée à la fin du XIX[e] siècle pour étudier les propriétés thermodynamiques des gaz avait conduit à attribuer à chaque résonateur à la température T une énergie égale à kT, où k était la constante de Boltzmann; l'énergie d'un cristal contenant 3N modes s'élèverait donc à

$$U = 3N\,kT.$$

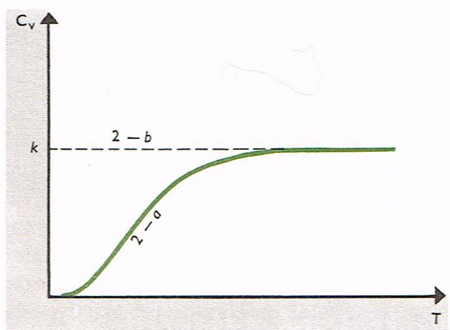
Il en résulte que la chaleur spécifique à volume constant, c'est-à-dire la quantité de chaleur qu'il faut fournir au cristal pour élever sa température d'un degré centigrade, devrait être égale à la dérivée de U par rapport à T :

$$C = \frac{\partial U}{\partial T} = 3\,Nk.$$

C'est la célèbre loi de Dulong et Petit, bien connue des collégiens, assignant à



L'énergie moyenne E d'un oscillateur quantique, suivant la statistique de Bose-Einstein, croît avec la température (courbe 1 *a*); la chaleur spécifique correspondante $C_v = \frac{dE}{dT}$ croît aussi (courbe 2*a*). Les courbes 1 — *b* et 2 — *b* en trait ponctué donnent le comportement prévu par la statistique classique et que l'expérience contredit.



tout solide une chaleur spécifique constante, indépendante de la température, égale à 6 calories par molécule-gramme. Cette loi, bien vérifiée à température ordinaire ou à haute température, est contredite par l'expérience à basse température : la chaleur spécifique de tous les solides baisse lorsque la température décroît, et tend vers zéro en même temps que la température absolue T.

Pour le rayonnement émis par un corps noir, une difficulté analogue avait déjà surgi, que Planck avait résolue en admettant que la lumière est émise par quanta ou photons, chacun d'eux ayant une énergie égale à $h\nu$, où ν est la fréquence lumineuse et h une constante, dite « constante de Planck ». De même, Einstein avait, vers la même époque, admis que les oscillations de chaque oscillateur d'un solide étaient quantifiées et que chaque quantum d'oscillation, appelé plus tard « phonon », avait une énergie $h\nu$, où ν était la fréquence propre de l'oscillation et h la constante de Planck. La statistique de Bose-Einstein avait prévu alors qu'un tel oscillateur, en équilibre thermodynamique à la température T, avait la probabilité

$$\frac{1}{\exp(h\nu/kT) - 1}$$

d'être excité, et, par conséquent, qu'il lui correspondait une énergie moyenne

$$E = \frac{h\nu}{\exp(h\nu/kT) - 1},$$

où k était la constante de Boltzmann et où le symbole $\exp(x)$ signifiait une exponentielle de x (e^x). A haute température $h\nu/kT$ est petit, et le terme exponentiel peut se développer :
$\exp h\nu/kT = 1 + h\nu/kT + \ldots$;
par suite l'énergie moyenne par oscillateur est $\dfrac{k\nu}{\left(1 + \dfrac{h\nu}{kT}\right) - 1} = kT.$

On retrouve le résultat classique dans le cas où justement l'expérience le confirme. Au contraire, à basse température $h\nu/kT$ est grand, le terme exponentiel au dénominateur est très grand, et l'énergie de chaque oscillateur devient beaucoup plus petite que kT. Le calcul donne effectivement pour les chaleurs spécifiques des différents solides des valeurs en très bon accord avec l'expérience.

On peut donc considérer un cristal comme un édifice parcouru en tout sens par des particules (les phonons ou quanta de vibration des ions du réseau); le nombre de ces phonons excités est d'autant plus grand que la température est plus élevée. Le concept de phonon s'est beaucoup développé ces dernières années et joue un rôle considérable dans presque tous les chapitres de la physique des solides. Par exemple, le transport des phonons de l'extrémité chaude d'une tige solide vers son extrémité froide permet d'expliquer la conductivité thermique des solides, ou, tout au moins, la part de cette conductivité qui n'est pas due aux électrons libres, mais aux mouvements thermiques des atomes d'un solide. Un autre exemple : comme nous le verrons plus loin, les électrons libres d'un conducteur sont responsables de la conductivité électrique de ce solide; en présence d'un champ électrique, les électrons libres reçoivent de l'énergie de ce champ électrique et la cèdent au réseau cristallin en émettant des phonons, c'est-à-dire en apportant de la chaleur au solide; on retrouve le dégagement de chaleur de l'effet Joule, associé comme on sait au passage du courant électrique dans un conducteur.

CRISTAUX


CUBIQUE
fluorite
grenats
QUADRATIQUE
mélinose
RHOMBOÉDRIQUE
smithsonite
apatite
HEXAGONAL
béryl
ORTHORHOMBIQUE
topaze
CLINORHOMBIQUE
crocoïte
rhodonite
TRICLINIQUE
AMORPHE
opale

les électrons dans les cristaux

Parmi les propriétés des solides, certaines appartiennent en propre aux atomes qui constituent les cristaux. Au contraire, d'autres propriétés dépendent de l'état de la matière, de la façon dont les atomes constituants sont disposés. Le fer par exemple n'a pas le même comportement électrique ou magnétique à l'état liquide qu'à l'état de cristal.

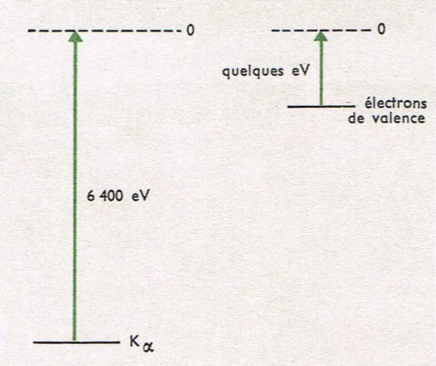
Il faut chercher à comprendre comment et pourquoi les atomes de fer peuvent se grouper pour former un cristal doué de propriétés nouvelles que ses atomes constituants ne possèdent pas séparément. La même question se pose déjà à propos des molécules, qui ont, nous le savons, des propriétés différentes de celles des atomes qu'elles contiennent. Le problème soulevé est lié à la nature de la liaison chimique : nous avions déjà souligné qu'un cristal n'était rien d'autre qu'une molécule géante.

aperçu sur la théorie des bandes, ou l'art de loger 1 000 000 000 000 000 000 000 d'électrons

Notre point de départ est l'atome, dont, nous l'avons vu plus haut, la mécanique quantique seule peut expliquer la structure et les propriétés. Rappelons qu'un atome est essentiellement formé d'un noyau minuscule où est concentrée presque toute la masse de la matière; ce noyau, qui porte une charge positive $+Ze$ (multiple entier de la charge élémentaire e), est entouré de Z électrons, particules beaucoup plus légères que le noyau et portant chacune une charge négative élémentaire $-e$; ces électrons gravitent autour d'un noyau sur des orbites stables plus ou moins grandes; l'ensemble, électriquement neutre, est un atome de « numéro atomique Z ». Pour arracher un électron à un tel atome, il faut dépenser une énergie minimale bien définie, appelée « énergie de liaison » de l'électron considéré. Nous pouvons donc admettre qu'un électron sur une orbite stable est dans un état d'énergie bien définie. Certains électrons situés sur des orbites « profondes », proches du noyau, sont très fortement liés au noyau : ils ont une énergie de liaison très grande, de l'ordre de quelques dizaines de milliers d'électron-volts. Les électrons situés sur des orbites plus larges, c'est-à-dire sur des couches moins « profondes », ont une énergie de liaison moins grande; les électrons situés sur les orbites les plus extérieures ont des énergies de liaison très faibles (quelques dizaines d'électron-volts). C'est par l'intermédiaire de ces électrons que les atomes peuvent interagir pour former des molécules, et ces électrons responsables des liaisons chimiques sont dits « électrons de valence »; les électrons des couches profondes ne jouent en pratique aucun rôle dans les liaisons chimiques, protégés qu'ils sont par les électrons de valence de la couche extérieure.

La mécanique quantique explique qualitativement et quantitativement cette architecture des atomes à partir des noyaux et des électrons; seule l'imperfection des méthodes de calcul limite l'accord entre les prévisions théoriques et l'infinie richesse des spectres atomiques. On peut de même espérer que la mécanique quantique expliquerait la formation de molécules et de cristaux à partir d'atomes; c'est ce que nous allons essayer de faire.

Supposons que nous ayons à notre disposition un grand nombre N d'atomes



Arracher un électron de l'orbite K_α du fer, orbite profonde, requiert une énergie d'environ 6 400 électron-volts; pour un électron de valence, l'énergie requise n'est que de quelques électron-volts.

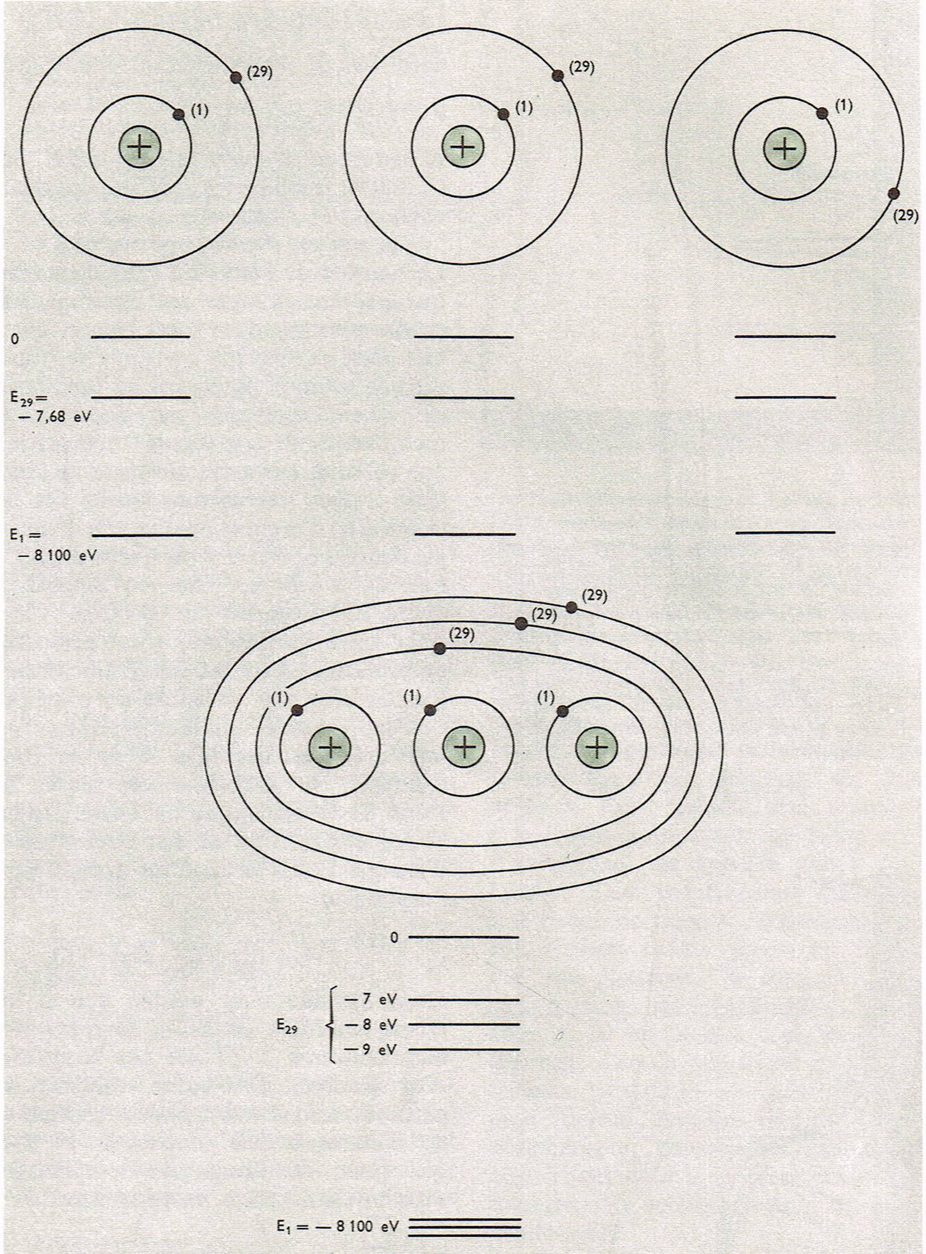
(atomes de cuivre, pour fixer les idées), situés à grande distance les uns des autres, de sorte que l'état de chacun d'entre eux ait une action tout à fait négligeable sur l'état des autres; les niveaux énergétiques sur lesquels sont situés les électrons de chaque atome ne sont donc pas modifiés et restent ceux d'un atome isolé. Rapprochons progressivement tous ces atomes; nous savons qu'une configuration stable est possible, puisqu'il existe un arrangement relatif de ces atomes qui constitue un cristal stable de cuivre. Lorsque la distance entre atomes décroît, chacun exerce sur les autres, et particulièrement sur ses voisins, une action croissante. Cette interaction est d'abord une attraction qui tend à rapprocher les atomes les uns des autres; si la distance décroît encore, des forces répulsives se manifestent et l'interaction devient répulsion. Il existe donc des distances interatomiques pour lesquelles ces interactions s'annulent, c'est-à-dire un arrangement de ces atomes qui représente un équilibre stable; comme nous l'avons montré plus haut, il en résulte en général un arrangement triplement périodique, c'est-à-dire un cristal. Lorsque nous rapprochons ainsi les atomes pour former un cristal, les niveaux énergétiques, d'abord inchangés, sont progressivement altérés.

Les niveaux des couches profondes, parce qu'ils correspondent à des électrons de grande énergie de liaison et qu'ils sont en outre protégés des interactions par les couches moins profondes, ne sont pratiquement pas modifiés; un électron d'un tel niveau ne « voit » pas les autres noyaux; il n'a d'yeux que pour le noyau auquel il est lié. Au contraire, les électrons de la couche extérieure, les électrons de valence, vont jouer un rôle fondamental dans l'interaction qui prend naissance; chaque électron de valence interagit fortement non seulement avec les autres électrons de valence, mais encore avec tous les ions de cuivre; il est désormais impossible d'affecter un électron à un atome déterminé de cuivre; **il faut considérer qu'un électron de valence est partout à la fois dans le cristal**; cette conception, typiquement quantique, ne doit pas surprendre le lecteur; après tout, la seule chose que l'on sache de cet

électron, c'est qu'il est à l'intérieur d'un cristal de cuivre de forme extérieure connue; il est donc naturel que l'état de cet électron soit défini par une fonction d'onde, nulle à l'extérieur du cristal et prenant des valeurs variées et non nulles à l'intérieur du cristal (au contraire, l'électron d'un niveau profond est défini par une fonction d'onde étroitement localisée autour d'un noyau donné de cuivre).

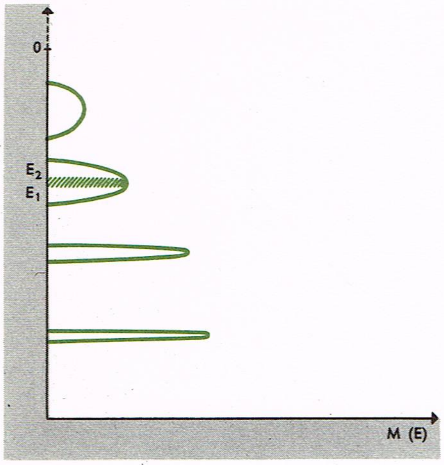
La figure ci-contre montre comment les niveaux d'énergie des électrons appartenant à trois atomes de cuivre se modifient lorsque les atomes se rapprochent pour former l'ébauche d'un cristal. Pour un grand nombre d'atomes, les phénomènes sont qualitativement analogues; en définitive, on peut prouver que tous les niveaux électroniques d'un cristal se groupent par bandes d'énergie : les électrons des couches profondes occupent des niveaux dont les énergies voisines forment des bandes très étroites, tandis que les électrons des couches extérieures, et en particulier les électrons de valence, occupent des niveaux groupés en bandes larges. **Toutes ces bandes d'énergie dites « bandes permises » sont séparées les unes des autres par des bandes d'énergie interdites.** Pour un cristal contenant N atomes (N est en général un nombre très grand), le nombre de niveaux d'une bande permise est en général un multiple de N; ce nombre, quoique fini, est si grand, qu'une bande permise peut être considérée comme continue, c'est-à-dire qu'à toute valeur d'énergie intérieure à cette bande correspond au moins un niveau possible pour un électron. Il est commode aussi d'introduire une « densité d'états » : si dN est le nombre de niveaux dont l'énergie est comprise entre deux valeurs E et E + dE voisines situées dans une bande permise, ce nombre, proportionnel à dE, peut s'écrire $dN = M(E)\,dE$, où le coefficient M, en général fonction de E, est la densité d'états.

Telle est, à peine esquissée, la célèbre théorie des bandes dans les solides. Grâce à elle nous avons résolu le problème suivant : nous savons quels états d'énergie les électrons d'un cristal donné peuvent occuper; ces états sont groupés en bandes d'énergie permises, séparées par des bandes d'énergie interdites. Cela conduit tout naturellement notre curiosité à poser la question suivante : **Comment les électrons vont-ils effectivement occuper ces places?**

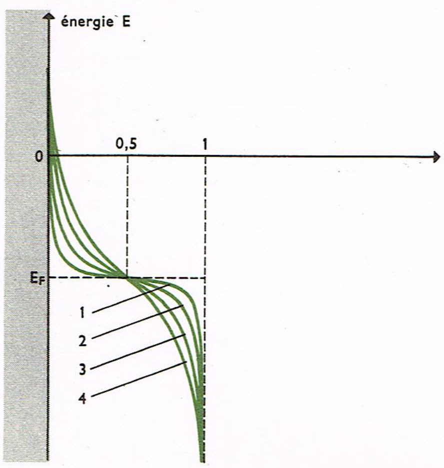


On a représenté trois atomes de cuivre, évidemment identiques, très éloignés les uns des autres; chaque atome possède 29 électrons, mais on a seulement fait figurer l'électron (1) occupant le niveau le plus profond (il décrit l'orbite la plus petite) et l'électron (29), électron de valence occupant un niveau peu profond (il décrit l'orbite la plus grande); entre ces deux orbites se trouvent celles de 27 électrons intermédiaires qui, pour raison de simplicité, ont été omises sur la figure. Les diagrammes indiquent les niveaux d'énergie de ces deux types d'électrons pour ces trois atomes de cuivre isolés. Lorsque ces trois atomes viennent au contact, formant l'ébauche d'un cristal de cuivre, les trois électrons (1) restent chacun localisé autour d'un noyau et occupent un niveau dont l'énergie est très voisine de l'énergie correspondante de l'atome isolé; au contraire, les électrons (29) communs aux trois atomes de cuivre occupent trois niveaux d'énergie très différents du niveau initial.

Remarque : les échelles de cette figure ne sont pas correctes; c'est ainsi que les trois niveaux E_1 du bas de la figure (voisins de — 8 100 eV) ne sont séparés les uns des autres que par des énergies extrêmement faibles, peut-être de l'ordre de 10^{-6} eV ! Les valeurs — 7, — 8 et — 9 eV sont de simples ordres de grandeur. Enfin il faut considérer les trajectoires des électrons comme des images très grossières n'ayant qu'une valeur descriptive.



Si l'on porte l'énergie verticalement en prenant pour origine l'énergie d'un électron en dehors du cristal et en abscisses la densité d'états M(E), on obtient le diagramme représenté ci-dessus; M(E) est évidemment nul dans les bandes d'énergie interdites et prend des valeurs non nulles dans les bandes d'énergie permises; l'aire hachurée est égale au nombre de niveaux dont l'énergie est comprise entre E_1 et E_2 [ce nombre d'ailleurs égal à l'intégrale $\int_{E_1}^{E_2} M(E)\, dE$].



On a l'habitude de porter sur un graphique l'énergie en ordonnées, suivant un axe vertical, et la probabilité d'occupation $f(E)$ d'un niveau électronique d'énergie E en abscisses, suivant un axe horizontal. A chaque température T correspond une courbe du graphique. (Les courbes 1, 2, 3, 4 correspondent à des températures croissant dans cet ordre.) On voit que pour toute température la probabilité d'occupation, égale à 1 pour les énergies les plus basses, devient égale à 1/2 pour l'énergie de Fermi et tend vers 0 pour les énergies les plus élevées; la probabilité f passe de un à zéro sur une « frange d'énergie » de quelques fois kT, l'énergie d'agitation thermique; ce saut est d'autant plus brusque que la température est plus basse.

l'occupation des bandes d'énergie d'un solide : un difficile problème de logement

Remarquons d'abord qu'un cristal de quelques millimètres cubes contient environ 10^{21} électrons
(1 000 000 000 000 000 000 000).
Le nombre de places ou états d'énergie que peuvent occuper ces électrons est encore plus grand, et il est bien évident que nous ne pouvons envisager de trouver une solution rigoureuse au problème de savoir comment ces places sont occupées. Seule une réponse approchée, une solution moyenne, présente de l'intérêt : nous chercherons quelle est la probabilité d'occupation d'un état d'énergie donné ou, si l'on veut, quelle proportion du temps cet état, en moyenne, abrite effectivement un électron. Calculer cette probabilité, c'est résoudre un problème de statistique. Si l'on admet que les électrons obéissent au principe de Pauli, principe qui leur interdit de se mettre à plusieurs dans le même état d'énergie, ce problème est résolu à l'aide de la statistique de Fermi-Dirac la probabilité f qu'un état électronique d'énergie E soit occupé est donnée par l'expression

$$f(E) = \frac{1}{1 + \exp[(E - E_F)/kT]}.$$

Cette formule n'est valable que si le cristal considéré est à une température bien uniforme T (T est une température absolue, c'est-à-dire comptée à partir du zéro absolu); k est la constante de Boltzmann déjà rencontrée, et E_F, une sorte de niveau de remplissage dont la signification va apparaître :

$$f(E_F) = \frac{1}{1 + \exp(0)} = \frac{1}{1+1} = 0{,}5.$$

En d'autres termes, E_F est l'énergie d'un état fictif qui a 50 chances sur 100 d'être occupé, et autant d'être inoccupé. Dès que E est supérieur à E_F de plusieurs fois l'énergie kT, le dénominateur de l'expression de f est grand devant 1 et la probabilité d'occupation devient très faible; au contraire, dès que E est inférieur à E_F de plusieurs fois kT, le dénominateur est très voisin de l'unité et la probabilité d'occupation devient très proche de un.

Considérons maintenant les bandes d'énergie de la figure précédente en fonction de la probabilité d'occupation; on voit immédiatement que les bandes les plus profondes sont occupées, puisque pour ces états la probabilité d'occupation est égale à 1 et que cette conclusion est vraie pour tous les états d'énergie nettement inférieure à E_F; au contraire, les états d'énergie nettement supérieure à E_F sont inoccupés. Ces résultats apparaissent sur la figure ci-contre, où les zones hachurées correspondent aux états occupés.

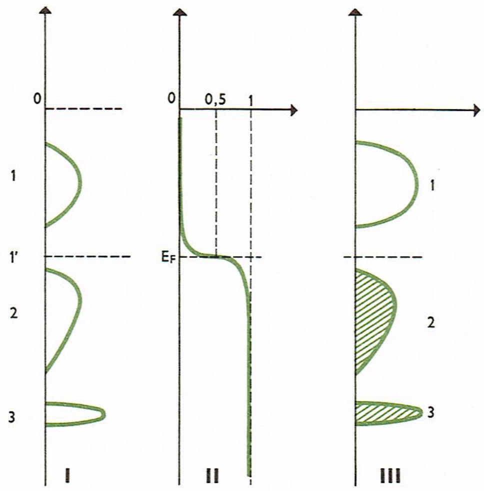
Deux cas peuvent se produire. Dans le premier (1), le niveau d'énergie de Fermi (E_F) tombe dans l'une des bandes interdites, de sorte que les bandes permises situées au-dessous de E_F sont complètement pleines et que les bandes permises situées au-dessus sont vides. Il n'y a donc que des bandes pleines ou des bandes vides; la bande pleine dont l'énergie est le plus élevée est dite « bande de valence »; la bande vide dont l'énergie est le plus basse est appelée « bande de conduction ».

Dans le deuxième (2), le niveau de Fermi tombe dans une bande d'énergie permise, de sorte que cette bande n'est jamais vide ni jamais pleine; toutes les autres bandes étant évidemment ou vides ou pleines.

Cette classification, qui permet de séparer les solides en deux classes, est d'une importance considérable, comme le montrent les remarques suivantes.

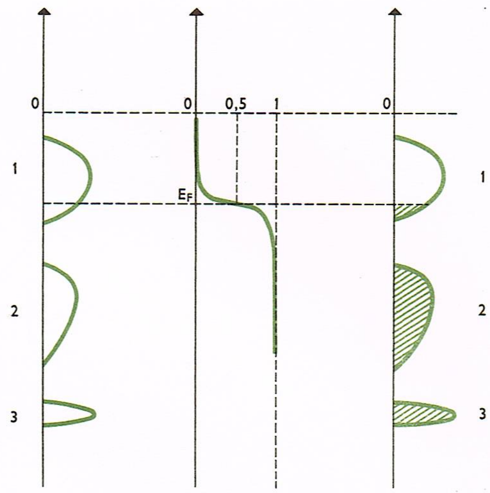
un problème de transport : conducteurs et isolants

S'il passe un courant électrique plus ou moins intense dans un solide auquel on applique un champ électrique, on dit qu'il s'agit d'un conducteur de l'électricité; tels sont par exemple les métaux. S'il ne passe pratiquement aucun courant électrique, on dit que le solide est un isolant; tel est par exemple le diamant. Le passage d'un courant électrique, c'est-à-dire le transport de charges électriques, peut provenir d'un déplacement des ions. C'est ainsi que le courant électrique dans un électrolyte est dû au mouvement des ions positifs et négatifs du liquide; ce mécanisme ne se produit qu'exceptionnellement dans les solides (l'exemple le plus connu est celui des cristaux ioniques). En général, le courant électrique dans un solide est dû au déplacement des électrons. En l'absence de champ électrique extérieur appliqué, aucun courant électrique n'existe, sinon le système dissiperait spontanément de l'énergie par effet



La partie gauche de la figure représente respectivement la densité des niveaux d'énergie d'un solide (I), la probabilité d'occupation (II) et l'occupation effective des niveaux (III); le niveau de Fermi E_F tombe dans une bande d'énergie interdite (I'); il en résulte que la bande (I) est vide, tandis que les bandes (2) et (3) sont occupées : on a affaire à un isolant ou à un semi-conducteur.

La partie droite de la figure est relative au cas d'un conducteur, un métal, par exemple; comme le niveau de Fermi tombe dans la bande d'énergie permise (I), celle-ci est partiellement remplie, alors que les bandes (2) et (3) sont complètement pleines.



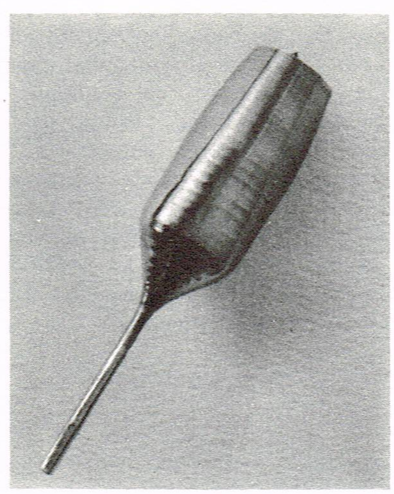
Joule, ce qui est absurde. Par suite, les électrons d'une bande partiellement ou complètement pleine ne transportent aucun courant en l'absence de champ électrique extérieur. Lorsqu'on applique un champ électrique, les électrons d'une bande pleine ne peuvent pas modifier leur mouvement pour créer un courant électrique, puisque toutes les places possibles sont occupées. Ce raisonnement serait en défaut si un électron pouvait sauter vers une place vide appartenant à une bande supérieure (on montre que ce mécanisme ne peut se produire que pour des champs électriques très intenses, 10^6V/cm par exemple; ce mécanisme est alors invoqué pour expliquer le « claquage » électrique des isolants). Deux électrons peuvent permuter en échangeant leurs places, mais, comme les électrons sont des particules indiscernables, cette permutation ne peut avoir d'effet et, en particulier, ne peut produire de courant; **donc, sous l'influence d'un champ électrique extérieur, seule une bande partiellement remplie donne naissance à un courant électrique.** Par conséquent, les solides de la première classe sont des isolants ou des semi-conducteurs (nous reviendrons plus loin sur ce point) et ceux de la seconde classe sont des conducteurs.

la liaison métallique

L'atome de sodium ne possède qu'un électron sur sa couche externe; comme un électron possède deux orientations possibles de spin, on peut montrer que pour un cristal de N atomes la bande de valence comporte 2N places pour N électrons et n'est qu'à demi remplie; nous pouvons prédire que le sodium est un conducteur de l'électricité, ce que l'expérience confirme. En d'autres termes, on a des ions sodium Na^+ noyés dans un nuage d'électrons négatifs; l'attraction des électrons sur les ions l'emporte sur la répulsion des ions entre eux et des électrons entre eux. Il en résulte une cohésion qui est propre à la liaison « métallique ». Le même raisonnement s'applique au cuivre, à l'argent, à l'or, etc. Et, de façon générale, à tous les métaux, quoique d'une manière parfois plus compliquée.

liaison de covalence et liaison ionique

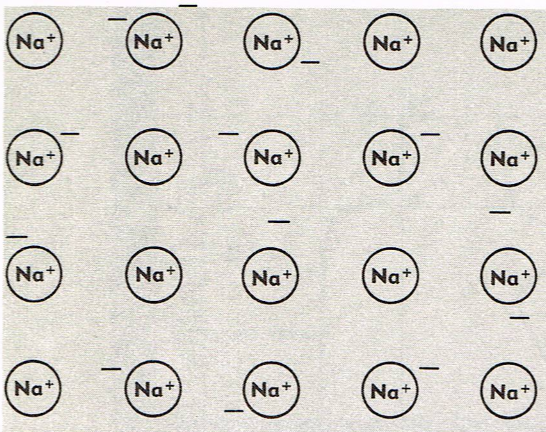
Le carbone, lui, possède quatre électrons sur sa couche de valence, et, lorsqu'il est cristallisé sous la forme de diamant (système cubique), chaque atome de carbone a quatre plus proches voisins formant un tétraèdre dont il est le centre; tous les atomes mettent en commun deux à deux une paire d'électrons, de sorte que chaque atome est entouré par huit électrons, quatre lui appartenant, quatre prêtés par ses quatre voisins; entre deux atomes voisins de carbone existe une sorte de pont électronique formé de deux électrons de spins opposés. Cette liaison, très semblable à la liaison responsable de la molécule d'hydrogène, est appelée une « covalence ». Tous ces électrons constituent une bande de valence entièrement saturée, puisque chaque atome est entouré par huit électrons : le diamant est un isolant ou un semi-conducteur. Le silicium et le germanium sont, comme le diamant, dans la quatrième colonne du tableau périodique et possèdent eux aussi quatre électrons sur la couche périphérique; comme le diamant, ils cristallisent dans le système cubique en donnant des solides isolants ou semi-conducteurs.

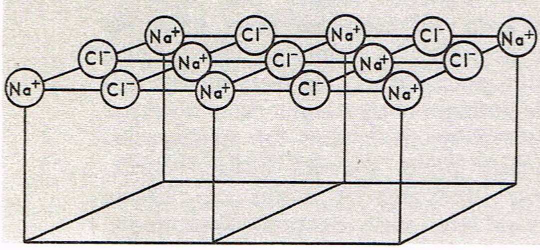


Le germanium est le plus parfait de tous les cristaux. Monocristal de germanium obtenu au C.N.E.T.

Pour les corps composés, la situation est plus compliquée. Une combinaison chimique dite « du type III-V », comme l'arséniure de gallium, est formée d'un métalloïde de la cinquième colonne possédant cinq électrons périphériques, l'arsenic, et d'un métal de la troisième colonne ayant trois électrons périphériques, le gallium; on voit que, là encore, chaque atome peut s'entourer de huit électrons complétant ainsi sa couche électronique; la bande de valence est pleine, et un tel cristal est un isolant ou un semi-conducteur. On arrive à la même conclusion avec les composés du type II-VI comme les oxydes ou les sulfures, et avec les composés du type I-VII comme les halogénures des métaux alcalins. Dans ce dernier cas

Doc. C. N. E. T.



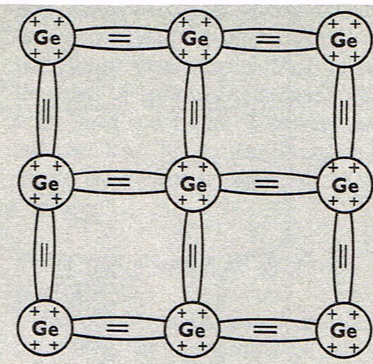


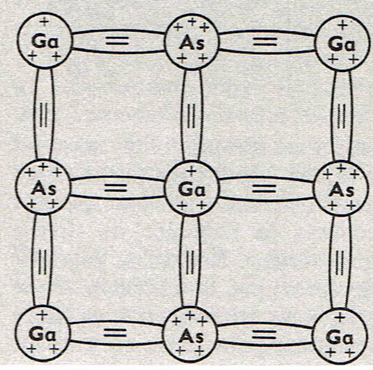
Les liaisons chimiques dans les cristaux. Lorsque l'on rapproche des atomes pour former un cristal, l'interaction des électrons des couches externes avec les ions conduit à des liaisons de différentes natures.

Dans le cas de la *liaison métallique* (le sodium, par exemple), les ions positifs sont noyés dans une mer d'électrons : ce gaz électronique joue le rôle d'un ciment liant les ions les uns aux autres et assure la cohésion du cristal (*fig. en haut, à gauche*).

Les *cristaux ioniques* sont un empilement d'ions alternativement positifs et négatifs; dans le chlorure de sodium, par exemple, chaque ion chlore Cl^- est entouré de 6 ions sodium Na^+, et réciproquement (*fig. en bas, à gauche*).

Dans un *cristal covalent*, une paire d'électrons forme entre deux plus proches voisins un pont électronique ; le germanium, formé d'ions Ge^{++++}, est purement covalent, alors que l'arséniure de gallium, formé d'ions Ga^{+++} et As^{+++++}, est covalent avec un certain degré d' « ionicité » (*fig. à droite*).





apparaît une nouvelle notion : **celle de liaison ionique**. Par exemple, dans les cristaux de chlorure de sodium on a, en quantités égales, des atomes de sodium qui n'ont qu'un électron sur leur couche externe et des atomes de chlore qui en ont sept. On pourrait imaginer que le chlore prête trois électrons au sodium, ce qui permettrait à chaque atome d'être, comme plus haut, entouré de huit électrons complétant une couche; en fait, les choses se passent différemment : le sodium donne son électron au chlore et devient alors un ion chargé positivement (d'une charge élémentaire $+ e$) dont la couche électronique extérieure est vide, tandis que le chlore devient un ion chargé négativement (d'une charge élémentaire $- e$) dont la couche électronique est pleine; ces deux ions de charges opposées ont l'un pour l'autre une attraction mutuelle bien compréhensible.

On dit qu'il s'agit d'une « liaison ionique »; on montre aussi que la bande de valence est pleine. Donc, tous les composés envisagés sont des isolants ou des semi-conducteurs; quand on parcourt la suite, depuis les cristaux de la colonne IV du tableau périodique jusqu'aux composés III-V, puis II-VI, et enfin I-VII, on passe progressivement de la liaison de covalence à la liaison ionique.

semi-conducteurs et isolants

Dans un solide appartenant à la classe des semi-conducteurs ou isolants, le niveau de Fermi est, par définition, à peu près au milieu d'une bande interdite. Si la largeur ΔE de cette bande interdite, c'est-à-dire la différence d'énergie entre le niveau le plus bas de la bande de conduction et le niveau le plus haut de la bande de valence, est beaucoup plus grande que kT (énergie d'agitation thermique), la probabilité d'occupation des niveaux de la bande de valence est pratiquement égale à un, et tous les niveaux de cette bande sont effectivement occupés. De même, la probabilité d'occupation des niveaux de la bande de conduction est nulle; aucun de ces niveaux n'est occupé. Supposons, par exemple, $\Delta E = 2{,}5$ eV; à température ordinaire, on a $kT \sim 0{,}025$ eV; en admettant que E_F est situé au milieu de la bande interdite, la probabilité d'occupation des états du bas de la bande de conduction est

$$f_c = \frac{1}{1 + \exp(\Delta E/2kT)} = \frac{1}{1 + \exp(50)} \simeq 10^{-22},$$

ce qui est un nombre extrêmement petit.

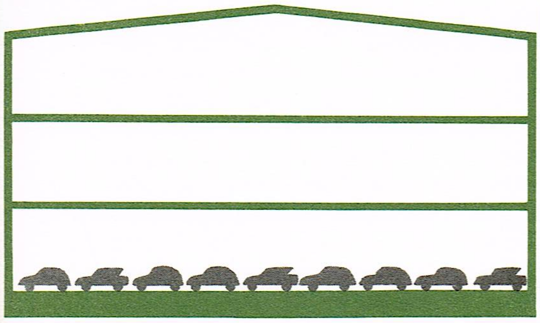
De même, la probabilité d'occupation des états du sommet de la bande de valence est

$$f_v = \frac{1}{1 + \exp(-50)} \simeq 1 - \exp(-50) = 1 - 10^{-22}.$$

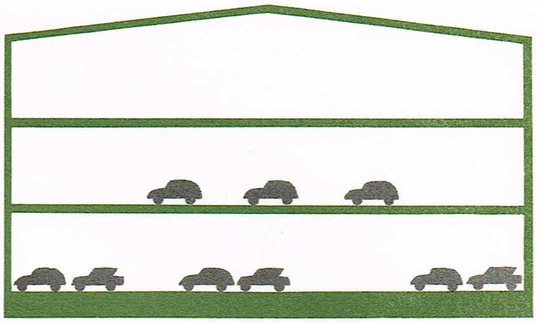
Dans ce cas, on dit que le solide est un isolant; aucun courant électrique ne peut y circuler.

Si le rapport $\Delta E/kT$ n'est pas aussi grand (soit parce que ΔE est plus petit, soit parce que la température est plus élevée), la probabilité d'occupation du fond de la bande de conduction n'est pas tout à fait négligeable et il existe dans cette bande quelques électrons quasi libres; de même, la probabilité d'occupation du sommet de la bande de valence n'est pas tout à fait égale à un; il y a donc au sommet de cette bande quelques états non occupés; dans ce cas, on dit que le solide est un semi-conducteur. **Il en résulte qu'à température assez basse tout semi-conducteur devient isolant et, qu'inversement, à température assez élevée, tout isolant devient semi-conducteur.**

On peut démontrer que ces quelques places vides se comportent en fait comme des particules quasi libres appelées « trous », de charge positive égale et de signe contraire à la charge des électrons et ayant à peu près la même masse; une image donnée par W. Shockley permet de visualiser assez bien cette notion de trous : la bande de valence est



De même que les automobiles d'un garage dont le rez-de-chaussée est plein et le premier étage vide sont condamnées à l'immobilité, de même les électrons d'un semi-conducteur à très basse température ne peuvent se déplacer.
Si quelques voitures montent au premier étage, elles peuvent se déplacer, autorisant du même coup le mouvement des voitures du rez-de-chaussée. De même, dans un cristal semi-conducteur, l'agitation thermique permet à quelques électrons de quitter la bande de valence en y laissant des trous : électrons et trous peuvent se déplacer.



analogue à l'étage inférieur d'un garage dont toutes les places, sauf quelques-unes, sont occupées; il est donc simple de décrire le mouvement de ces voitures en décrivant le déplacement des places vides; la bande de conduction, elle, est analogue à l'étage supérieur, qui, à peu près vide, ne contient que quelques voitures. A très basse température, l'étage supérieur est tout à fait vide, et l'étage inférieur est plein, sans une place vide; en d'autres termes, l'agitation thermique a pour effet de faire monter quelques voitures de l'étage inférieur à l'étage supérieur. Electrons et trous contribuent à la conduction électrique; les uns et les autres sont en nombre d'autant plus grand que la température est plus élevée.

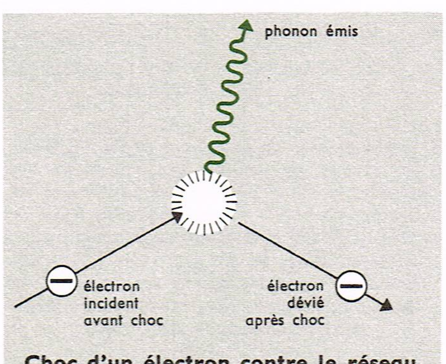
Tous les raisonnements faits plus haut reposent sur une hypothèse qu'il est bon de souligner maintenant : **les bandes d'énergie décrivent les états possibles d'un électron et d'un seul;** il semblerait qu'en tenant compte de l'occupation des niveaux par les autres électrons on va modifier les états d'énergie de l'électron considéré; en fait, on admet que les bandes d'énergie relatives à un électron tiennent compte de la façon dont tous les autres électrons occupent les niveaux restants. Le fait que cela soit possible résulte d'une approximation, dite « approximation à un électron »; elle est en général tout à fait légitime.

Dans ce chapitre, les faits ont été énoncés sous une forme aussi simple que possible en essayant d'en montrer l'aspect à la fois logique et intuitif. L'appareil mathématique de la mécanique quantique, évidemment exclu d'un tel exposé, permet d'établir de façon précise tous ces résultats.

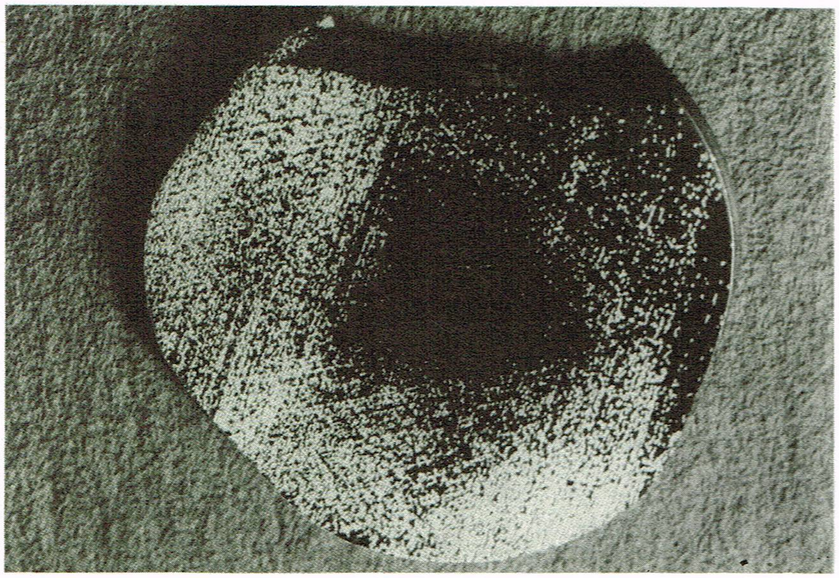
Résumons-nous. On peut, avec une bonne approximation, considérer qu'un électron dans un cristal est dans un état d'énergie appartenant à une bande d'états permis. D'où résulte la distinction entre conducteurs et non-conducteurs. Pour ces derniers, on définit une bande d'énergie interdite ΔE; si ΔE est très grand par rapport à l'énergie d'agitation thermique, on a un véritable isolant; dans le cas contraire, on a un semi-conducteur. Enfin, la notion de cohésion cristalline permet de définir trois types de liaison : métallique, covalente, ionique (**on a omis la liaison de type moléculaire, dont l'intérêt pratique est négligeable**).

les semi-conducteurs

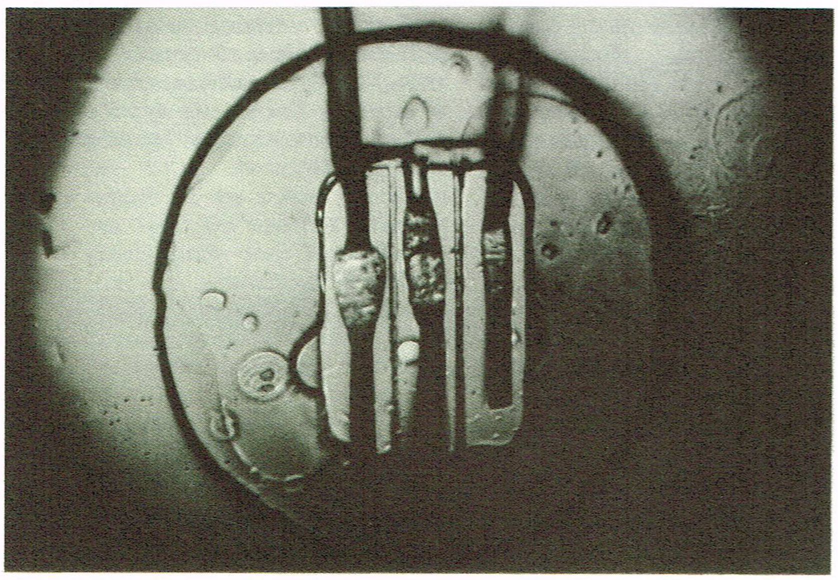
Nous avons vu qu'un semi-conducteur est un solide dont, à basse température, la bande de valence est pleine et la bande de conduction vide, les deux bandes étant séparées par une bande d'énergie interdite ΔE. A température pas trop basse, l'agitation thermique est suffisante pour que quelques électrons du sommet de la bande de valence viennent occuper les places du fond de la bande de conduction. Ces électrons de la bande de conduction se comportent comme des particules « presque » libres : sous l'influence d'une force extérieure constante, un champ électrique par exemple, ils acquièrent un mouvement uniformément accéléré comme le feraient des particules de charge négative $-e$ et de masse m_e constante. (Nous nous plaçons dans une approximation non relativiste, justifiée par la petitesse, en comparaison de la vitesse de la lumière, des vitesses mises en jeu.) Ces électrons ne sont pas « complètement » libres : en effet, leur masse m_e, dite « masse effective », diffère en général d'un semi-conducteur à un autre et n'est pas égale à la masse de l'électron dans le vide; de plus, le mouvement uniformément accéléré que prend un électron d'un semi-conducteur dans un champ électrique constant ne saurait se prolonger très longtemps, puisque cet électron interagit au bout d'un temps plus ou moins long soit avec un autre électron, soit avec une impureté ou un défaut de réseau, soit encore avec les phonons ou quanta de vibration du réseau cristallin. Précisons ce dernier cas. Un électron dans un cristal est décrit par une fonction d'onde se propageant sans affaiblissement dans le potentiel périodique du cristal; cela n'est pas possible pour n'importe quel état d'énergie. La preuve en est qu'il existe des énergies interdites; pour les valeurs d'énergie permises, cette propagation n'est possible que parce que le potentiel cristallin dans



Choc d'un électron contre le réseau cristallin avec émission d'un phonon.



En même temps que progressent nos connaissances fondamentales des solides, la technologie des semi-conducteurs se perfectionne sans cesse. Pour découper les cristaux de germanium ou de silicium, pour les roder, les polir, les façonner, pour traiter leurs surfaces, l'ingéniosité humaine fait des merveilles. La photographie supérieure montre des dislocations révélées par un polissage chimique; pourtant, une dislocation n'est qu'une simple rangée manquante d'atomes! La photographie inférieure montre un dispositif encore à l'étude au laboratoire : un transistor à « effet de champ », dispositif qui, dans quelques années, concurrencera peut-être les transistors classiques.



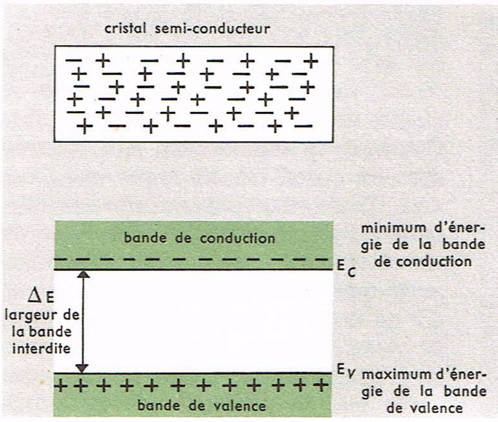


Diagramme d'énergie d'un semi-conducteur.

On a porté, en fonction de la distance le long d'un cristal, les énergies des électrons dans un semi-conducteur; les électrons et les trous présents dans le cristal *(dans le haut de la figure)* apparaissent dans le diagramme d'énergie *(en bas)* : les électrons au fond de la bande de conduction, les trous au sommet de la bande de valence.

lequel se meut l'électron est parfaitement périodique. S'il existe des phonons, c'est-à-dire si certains modes de vibration du cristal sont excités, le potentiel instantané dû aux ions n'est plus tout à fait périodique; cela trouble la propagation de l'onde pilote de l'électron au point de lui faire modifier brusquement son mouvement. En d'autres termes, l'électron subit un « choc » contre les vibrations du réseau; on peut montrer qu'au cours du choc un phonon est émis ou absorbé.

De façon tout à fait analogue, les trous, ou manques d'électrons dans le sommet de la bande de valence, peuvent être considérés comme des particules « presque libres » de masse m_t et de charge positive $+e$.

Electrons et trous sont en nombre identique, puisqu'un trou n'existe que parce qu'un électron de la bande de valence a quitté sa place pour venir occuper un état de la bande de conduction; on dit qu'il y a des **« paires électron-trou »** en nombre n_i par unité de volume; leur nombre doit du reste augmenter très rapidement avec la température. On montre théoriquement, et l'expérience le confirme, que n_i est donné par la formule $n_i = AT^{3/2} \exp - (\Delta E / 2kT)$ [où A est une constante, T la température absolue, k la constante de Boltzmann et ΔE la largeur de la bande interdite]. Sous l'influence d'un champ électrique, les électrons se déplacent dans une direction opposée à celle du champ électrique en transportant des charges négatives, tandis que les trous se déplacent dans la direction du champ en transportant des charges positives; ces deux contributions s'ajoutent, et l'on peut dire que la conductivité électrique d'un semi-conducteur est la somme de deux contributions : celle des électrons de la bande de conduction, celle des trous de la bande de valence.

les donneurs et les accepteurs

Bien avant que la notion de semi-conductibilité ne se fût dégagée, on avait déjà remarqué que des cristaux comme le germanium, le silicium, l'oxyde de cuivre et bien d'autres avaient des conductivités électriques variables d'un échantillon à l'autre. Prenons comme exemple le germanium, un semi-conducteur typique, qui est aussi le solide le mieux connu à l'heure actuelle; supposons que, dans un tel cristal, un atome de germanium soit remplacé par un atome d'arsenic

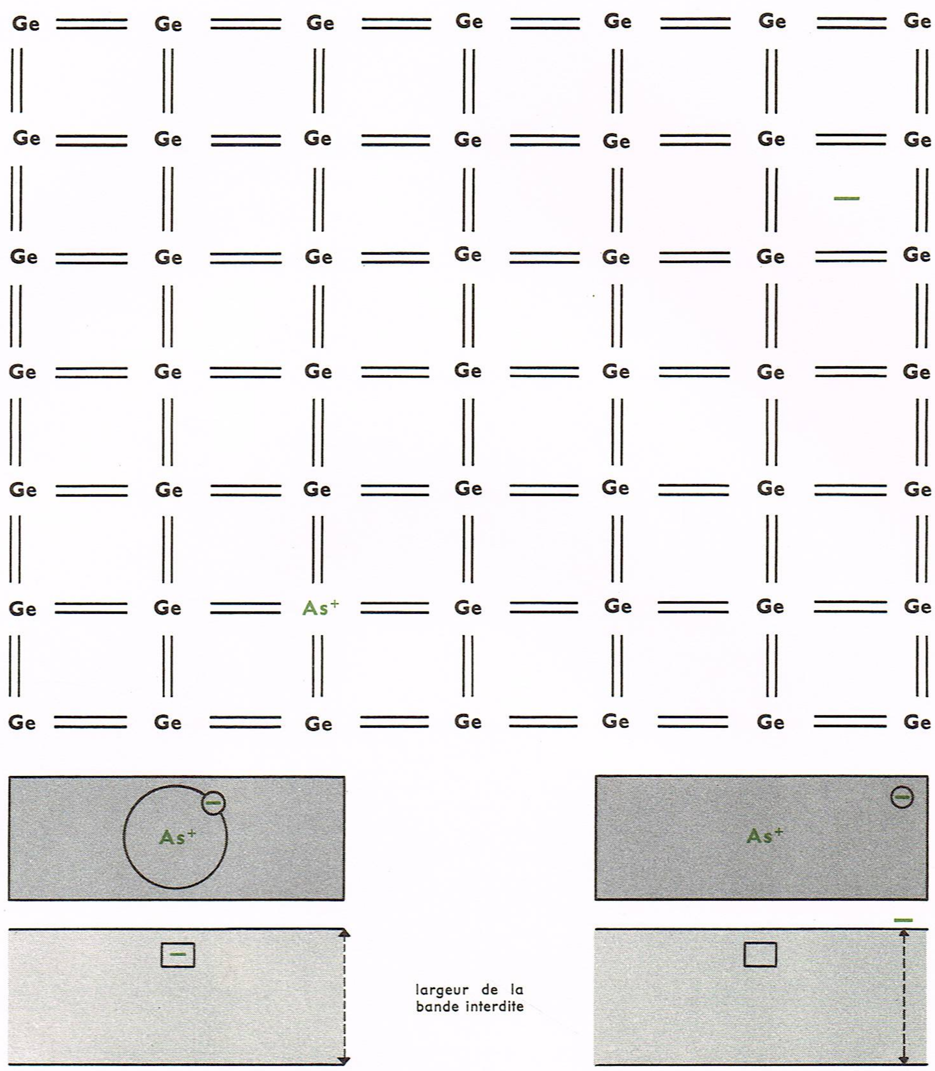
Doc. C. N. E. T.

(très semblable à l'atome de germanium, mais possédant sur sa couche externe cinq électrons au lieu de quatre); on peut penser que quatre électrons vont s'engager dans des covalences avec les quatre atomes de germanium voisins, le cinquième électron étant en excès. Admettons que cet électron soit libre, loin de l'atome d'arsenic, et occupe un état du fond de la bande de conduction; sur cet atome apparaît donc une charge positive non compensée, créant dans son voisinage un potentiel électrique attractif pour les charges négatives : l'électron supplémentaire peut se lier à l'atome d'arsenic de la même manière qu'un électron se lie à un proton pour former un atome d'hydrogène; dans ce dernier cas, l'énergie de liaison est égale à 13,6 électron-volts, alors que, pour l'électron lié à l'atome d'arsenic dans un cristal de germanium, l'énergie de liaison est d'environ 0,01 électron-volt, soit 1500 fois plus faible, principalement parce que la constante diélectrique du germanium est très élevée ($\varepsilon = 16$). L'électron ainsi lié à l'arsenic est donc sur un niveau d'énergie situé à 0,01 eV au-dessous de l'énergie qu'il a lorsqu'il est libre, c'est-à-dire au fond de la bande de conduction; on peut dire que l'arsenic crée dans le cristal de germanium un niveau d' « énergie pour électron », situé à 0,01 eV au-dessous du bas de la bande de conduction. A très basse température, l'électron occupe ce niveau et reste localisé au voisinage de l'atome d'arsenic; mais, lorsque la température n'est plus négligeable devant l'énergie de liaison (0,01 eV), l'électron est excité par l'agitation thermique et devient un électron libre du bas de la bande de conduction; tout se passe comme si l'atome d'arsenic avait « donné » un électron; on dit que c'est une impureté « donatrice » ou, plus simplement, un atome « donneur ». Un cristal de germanium contenant des traces d'arsenic possède, à température pas trop basse, autant d'électrons que d'atomes d'arsenic; on montre d'ailleurs que le nombre de trous est fortement réduit; plus précisément, si l'on désigne par N_D le nombre d'atomes d'arsenic par unité de volume et par n et p respectivement les nombres d'électrons et de trous par unité de volume, on a $p << n_i << n$ et $n \simeq N_D$.

Les autres atomes de la colonne V du tableau périodique (phosphore, antimoine) se comportent comme l'arsenic et sont aussi des donneurs. On peut refaire les raisonnements ci-dessus en remplaçant un atome de germanium, non plus par un atome d'arsenic, mais par un atome de gallium : le gallium ayant trois électrons sur sa couche externe, il existe un manque d'électron ou trou; il en résulte un niveau donneur de trou (ou accepteur d'électron) situé à 0,01 électron-volt au dessus du haut de la bande de valence; ce trou, libéré par agitation thermique, vient occuper un niveau du haut de la bande de valence; si l'on désigne par N_A le nombre d'atomes de gallium par unité de volume, on a $n << n_i << p$ et $p \simeq N_A$.

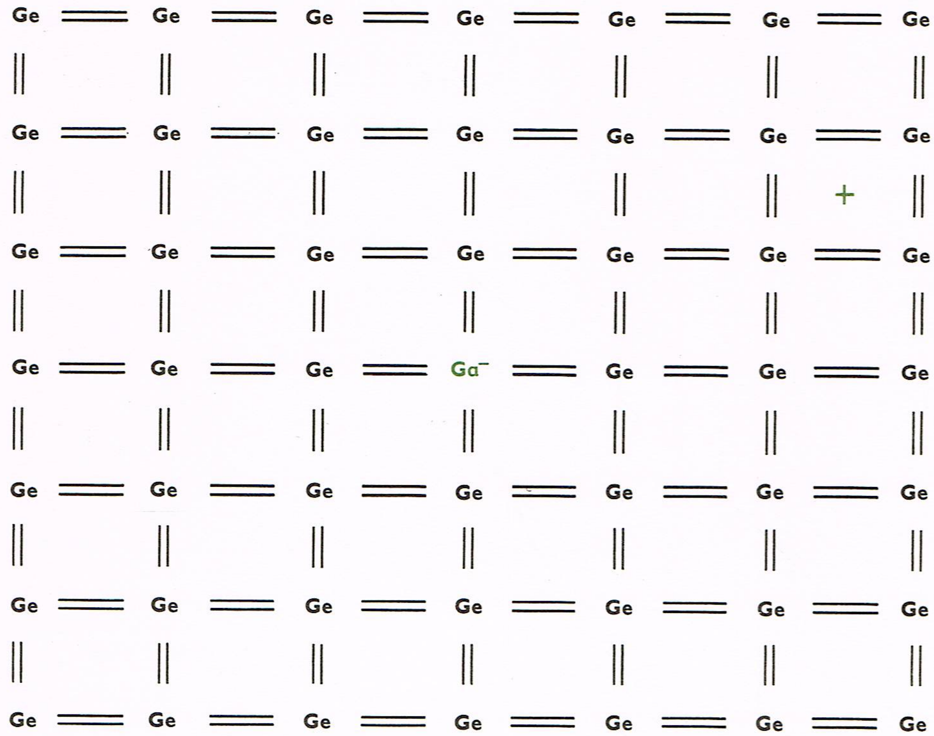
Le gallium est une impureté acceptrice ou un atome accepteur; les autres éléments de la colonne III du tableau périodique, bore, aluminium, indium, sont aussi des accepteurs.

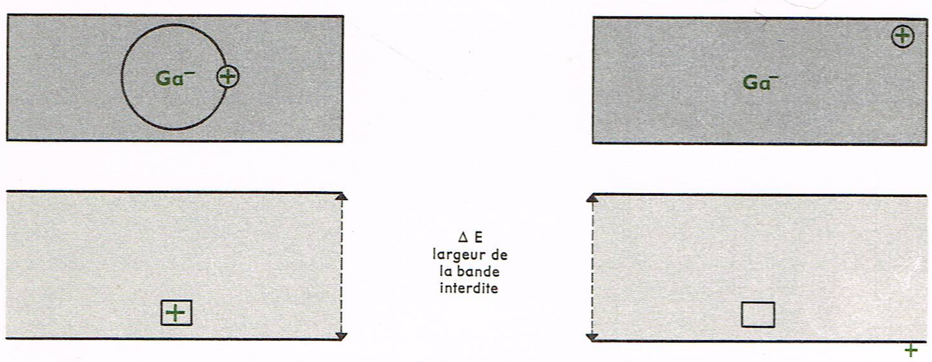
Il en résulte des conséquences très importantes. Tout d'abord, si un cristal de germanium contient des traces d'impuretés du type donneur, arsenic par exemple, sa conductivité électrique se fait essentiellement par électrons, le nombre de trous étant faible; on dit que le cristal est du **type *n***, et sa conductivité est proportionnelle à la concentration des donneurs. De même, si un cristal de germanium contient des traces d'impuretés du type accepteur, gallium par exemple, sa conductivité électrique se fait essentiellement par trous, le nombre d'électrons étant faible; on dit que le cristal est du **type *p***, et sa conductivité



Les impuretés dans les semi-conducteurs.

L'arsenic dans le germanium introduit un niveau donneur dans la bande interdite : à basse température *(partie gauche de la figure)*, l'électron décrit une orbite autour de l'ion arsenic (As^+) en occupant le niveau donneur; à température plus élevée *(partie droite de la figure)*, l'agitation thermique libère cet électron, qui reprend sa course vagabonde libre dans le cristal.





Les impuretés dans les semi-conducteurs.

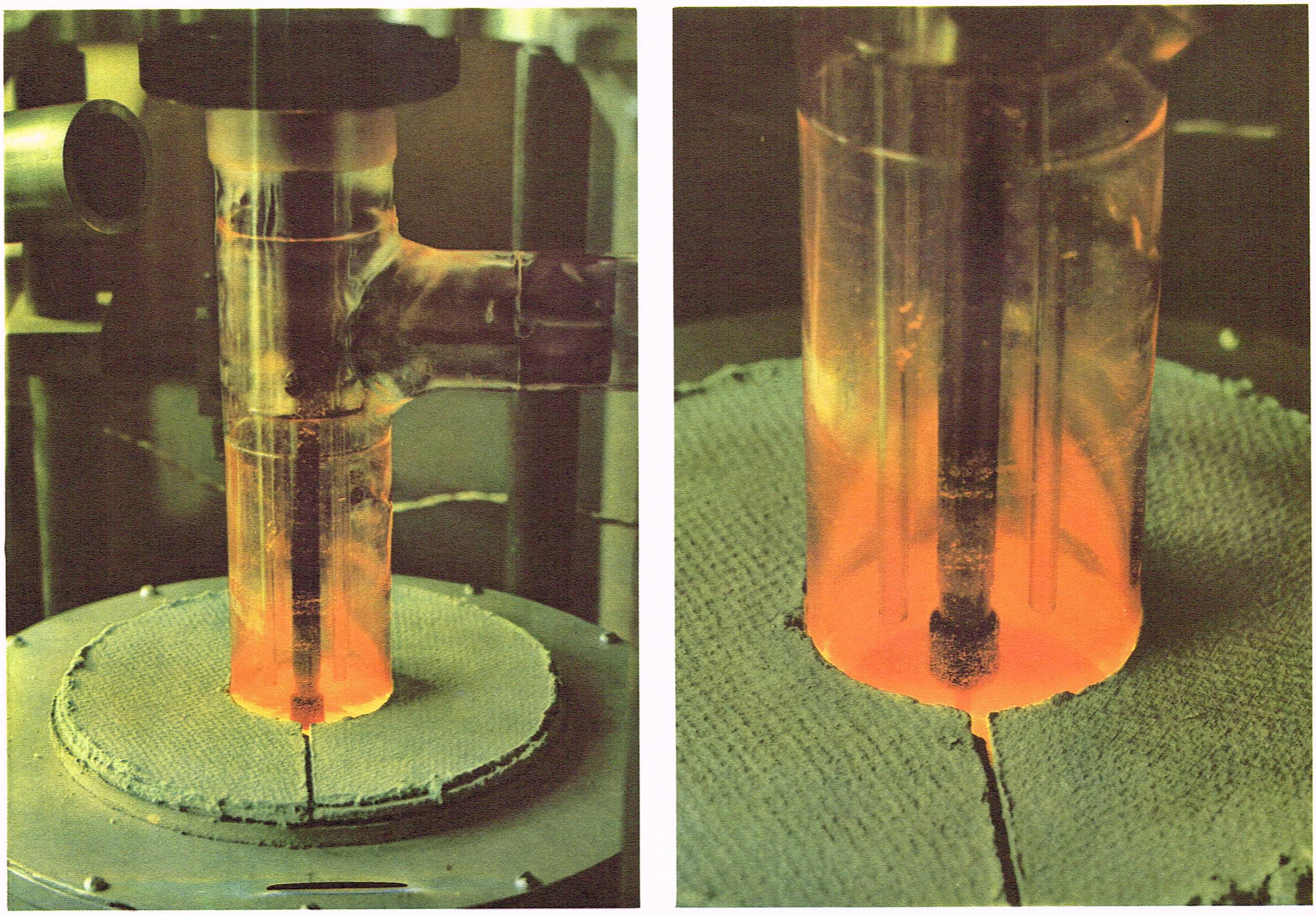
Un atome de gallium dans un cristal de germanium introduit un niveau accepteur dans la bande interdite : à basse température *(partie gauche de la figure)*, le trou, ou manque d'électron, décrit une orbite autour de l'ion de gallium (Ga^-) en occupant le niveau accepteur; à température plus élevée *(partie droite de la figure)*, l'agitation thermique libère ce trou, qui reprend sa course vagabonde libre dans le cristal.

est proportionnelle à la concentration des accepteurs. On comprend maintenant pourquoi les cristaux relativement impurs ont des conductivités variables d'un échantillon à l'autre selon la concentration des impuretés.

Pour bien comprendre les propriétés d'un semi-conducteur, il faut préparer des cristaux de grande pureté et même, plus généralement, de grande perfection, c'est-à-dire contenant aussi peu de défauts que possible. C'est la tâche à laquelle s'emploient dès 1944 les chercheurs des laboratoires de la Bell Telephone Company lorsqu'ils commencent à cette époque d'étudier les propriétés semi-conductrices du germanium et du silicium; ils découvrent des méthodes de purification d'une efficacité extrême, mesurent sur toutes leurs coutures des milliers d'échantillons et, en quelques années, obtiennent des cristaux dont la pureté dépasse 10^{-10}, c'est-à-dire contenant moins d'un atome étranger pour 10 000 000 000 d'atomes. En préparant ensuite des échantillons de germanium et de silicium volontairement contaminés avec les impuretés mentionnées plus haut (phosphore, arsenic, antimoine, bore, aluminium, gallium, indium), ils vérifient que les impuretés jouent un rôle déterminant dans les propriétés électriques des semi-conducteurs, en accord avec la théorie esquissée plus haut.

La notion de pureté chimique et de perfection cristalline, le rôle joué par impuretés et défauts dans les solides, ignorés jusqu'alors, apparaissent d'un seul coup en pleine clarté. Si certaines propriétés des solides sont essentiellement fonction de la quantité de matière mise en jeu, il devient évident que d'autres propriétés dépendent de traces infimes de certaines impuretés spécifiques. La même idée se fait jour dans différents domaines : certains matériaux nucléaires, le graphite par exemple, exigent eux aussi une grande pureté. Cette prise de conscience du rôle joué par les imperfections en général, et les impuretés en particulier, n'a pas une portée philosophique très considérable, mais elle a joué un rôle fondamental dans la plupart des progrès de la recherche technique au cours des quinze dernières années.

Les considérations développées plus haut sur les impuretés, sur les électrons et les trous, sont qualitatives et incomplètes; on peut traiter ces problèmes d'une façon approfondie à l'aide de la théorie des bandes et prévoir par le calcul les propriétés électroniques des semi-conducteurs; les résultats expérimentaux, en général, sont en excellent accord avec les prévisions théoriques. Non seulement le développement prodigieux de la physique des semi-conducteurs prouve la puissance explicative de la théorie moderne des solides fondée sur la mécanique quantique, mais encore des applications révolutionnaires se font jour. Vers 1945-1947, W. Shockley, J. Bardeen, W. Brattain étudient aux laboratoires Bell le mouvement des électrons et des trous dans le germanium, analysent leur comportement et montrent comment ces particules interagissent les unes sur les autres; ils découvrent que l'on peut, grâce à ces particules, injecter un signal électrique dans un cristal semi-conducteur et le faire ressortir amplifié : c'est le **transistor à pointe.** Puis, voilà en 1949 la **jonction *p - n* :** avec un cristal semi-conducteur dont une région est de type *n*, et l'autre de type *p*, on obtient un redresseur d'une incomparable qualité. Enfin, un **cristal *p - n - p*** constitue le transistor à jonction, qui supplante le transistor à pointe et, en quelques années, remplace les tubes électroniques dans de nombreuses applications.

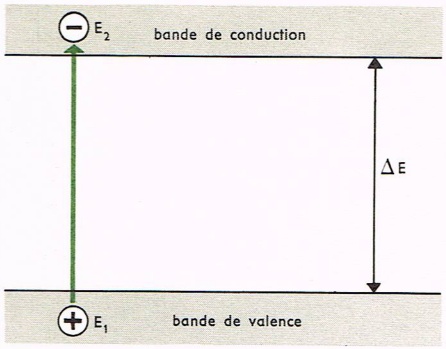


Deux phases de la croissance d'un cristal de germanium.

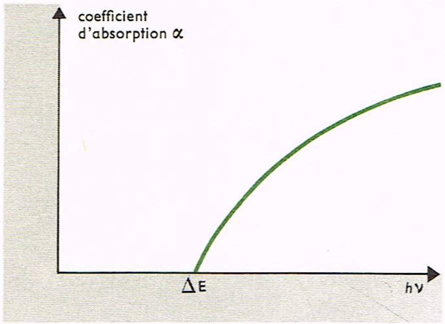


Fluorescence de divers cristaux éclairés en lumière ultraviolette

Phot. Laboratoire du Centre national d'études des télécommunications.



Un photon d'énergie $h\nu = E_2 - E_1$ est absorbé en excitant un électron de la bande de valence vers la bande de conduction, faisant ainsi apparaître une paire électron-trou.



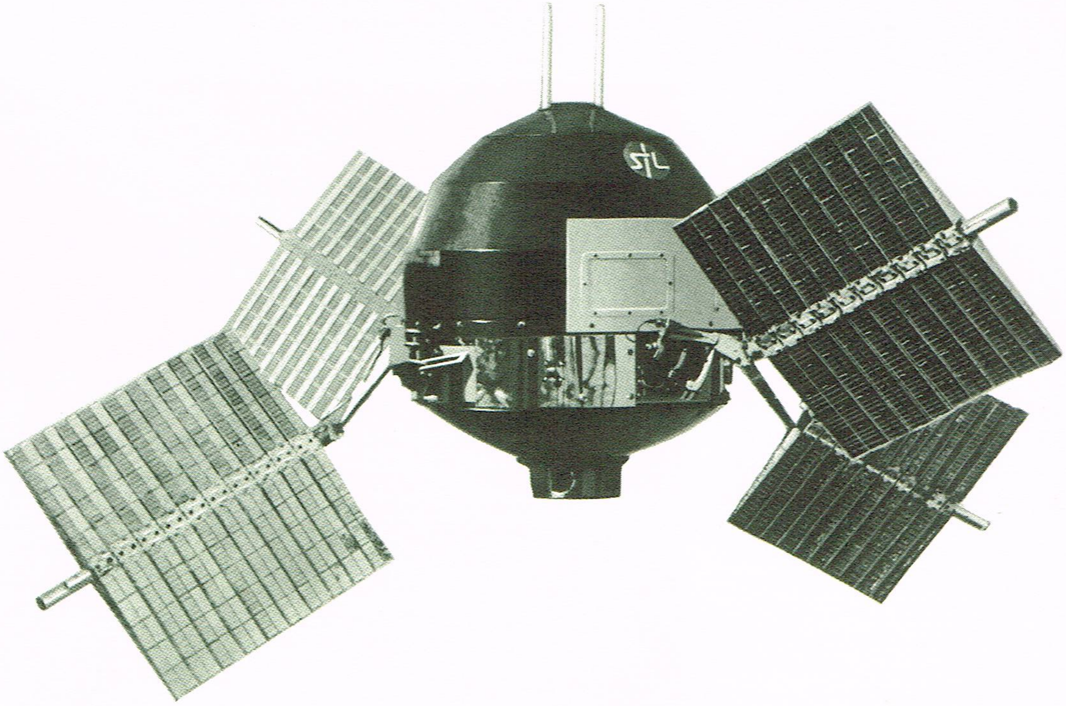
Courbe du coefficient d'absorption optique d'un semi-conducteur en fonction de l'énergie $h\nu$ des photons.

les mille feux du diamant et l'éclat métallique du germanium

On sait qu'un rayonnement électromagnétique de fréquence ν appartenant au spectre visible, ultraviolet ou infrarouge, peut être absorbé par un atome; dans un tel processus, un électron d'un niveau d'énergie E_1 est porté dans un état excité d'énergie E_2, tandis qu'un photon ou quantum d'énergie électromagnétique d'énergie $h\nu$ est absorbé, prélevé au rayonnement excitateur; la conservation de l'énergie entraîne $h\nu = E_2 - E_1$. L'absorption se produit à une fréquence bien déterminée; on dit que le spectre d'absorption présente des raies (le processus inverse est possible : réémission d'un photon de même fréquence $h\nu$, tandis que l'électron revient du niveau excité E_2 au niveau de base E_1; le spectre d'émission est encore un spectre discret, un spectre de raies).

Dans un solide, un semi-conducteur par exemple, les phénomènes sont un peu différents : un électron occupant un état d'énergie E_1 de la bande de valence peut être porté dans un état d'énergie E_2 de la bande de conduction en absorbant un photon d'énergie $h\nu = E_2 - E_1$. Si l'énergie $h\nu$ est plus petite que la largeur ΔE de la bande interdite, aucune absorption n'est possible et le semi-conducteur est transparent pour le rayonnement; au contraire, pour tous les photons d'énergie $h\nu$ supérieure à ΔE, l'absorption est possible et le semi-conducteur est opaque pour ces rayonnements; on a donc une bande d'absorption présentant un seuil d'énergie $h\nu = \Delta E$. L'expérience permet en général de tracer la courbe donnant le coefficient d'absorption en fonction de l'énergie. On a ainsi un moyen de déterminer la largeur ΔE de la bande interdite. C'est ainsi que le germanium, ayant une bande interdite $\Delta E = 0{,}70$ eV, est transparent pour les longueurs d'onde plus grandes que 1,8 μ, et absorbant pour les rayonnements de longueurs d'onde plus courtes que 1,8 μ (et en particulier pour le rayonnement visible). A cet égard, il doit se comporter comme un métal, et, en effet, le germanium présente l'éclat métallique. Au contraire, la largeur de la bande interdite du diamant est de 7 eV; ce cristal doit être

Cet étrange « moulin à vent » est le satellite « Explorer VI », photographié au cours d'essais portant sur ses batteries solaires à semi-conducteurs.



Phot. U. S. T. S.

transparent pour les rayonnements de longueurs d'onde supérieures à 0,18 μ, et, en particulier, pour la lumière visible; le diamant doit être transparent et incolore. Comme l'indice de réfraction y est très élevé, la lumière du jour qui y pénètre n'en ressort qu'après de multiples réflexions totales, d'où l'éclat du diamant et les mille feux qu'il jette pour séduire nos compagnes; comme on le voit, la physique des solides explique bien des choses!

les piles solaires des satellites

Lorsqu'un photon d'énergie $h\nu$ supérieure à ΔE est absorbé dans un semi-conducteur, nous avons vu qu'un électron de la bande de valence est porté dans la bande de conduction; il apparaît donc un trou dans la bande de valence et un électron dans la bande de conduction; les paires électron-trou créées par l'absorption du rayonnement produisent une conductivité supplémentaire permettant de détecter la présence de ce rayonnement. Tel est le principe de la photoconductivité, et l'origine de très nombreuses applications. En utilisant une jonction semi-conductrice *n - p*, on peut convertir l'énergie lumineuse en énergie électrique; avec le silicium, dont la bande interdite $\Delta E = 1{,}15$ eV correspond au proche infrarouge, on peut construire les batteries solaires qui équipent les satellites. Avec un semi-conducteur de bande interdite plus petite, comme l'antimoniure d'indium avec $\Delta E = 0{,}23$ eV, la sensibilité spectrale s'étend jusque vers 6 μ : on réalise ainsi des détecteurs à infrarouge aux performances très poussées.

luminescence et phosphorescence

Que deviennent les paires électron-trou produites par un rayonnement excitateur? Il faut que d'une façon ou d'une autre, en un temps plus ou moins long, l'électron excité occupant l'état d'énergie E_2 revienne à l'état d'énergie E_1 de la bande de valence qu'il occupait initialement. En d'autres termes, il faut que l'électron et le trou s'annihilent ou, comme on dit, se recombinent en restituant l'énergie $h\nu = E_2 - E_1$ prise au rayonnement excitateur. Cette recombinaison peut se faire de bien des façons. Le cas le plus simple est celui où se produit le processus inverse du processus d'excitation : il y a réémission d'un photon d'énergie $E_2 - E_1$; le cristal est luminescent et émet un rayonnement d'énergie égale à ΔE, largeur de la bande interdite, ou un peu supérieure. Il peut aussi arriver que l'électron revienne à sa place initiale dans la bande de valence en deux étapes : dans un premier temps, il tombe sur un niveau E_i de la bande interdite, créé par une impureté ou par un défaut, en émettant un photon d'énergie $h\nu_2 = E_2 - E_i$; dans un deuxième temps, l'électron passe de l'état E_i à l'état E_1 de la bande de valence ou, ce qui revient au même, le trou monte du niveau E_1 sur le niveau E_i : il y a alors émission d'un deuxième photon d'énergie $h\nu_1 = E_i - E_1$; le cristal est luminescent et émet des rayonnements d'énergie $h\nu_2$ et $h\nu_1$ (inférieure à ΔE, c'est-à-dire de longueur d'onde supérieure à celle du rayonnement excitateur).

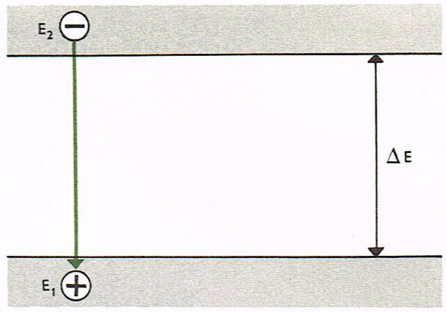
Il peut d'ailleurs arriver que l'une des deux transitions se produise sans émission de photons : l'énergie correspondante, par exemple $E_2 - E_i$, apparaît sous forme de phonons ou quanta de vibration du réseau cristallin, c'est-à-dire, en fait, sous forme de chaleur; on dit que la transition $E_2 \rightarrow E_i$ est non radiative.

Il arrive aussi parfois que l'une des deux transitions se produise très lentement, de sorte que le cristal émet de la lumière longtemps après avoir été éclairé par le rayonnement excitateur; c'est le phénomène de la phosphorescence.

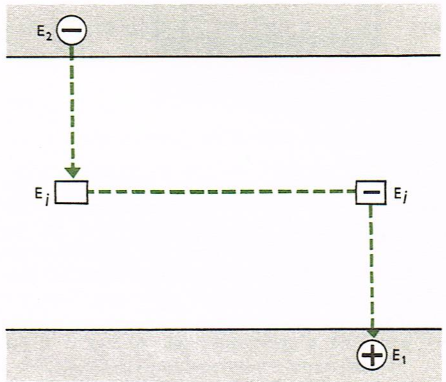
Des électrons et des trous « excédentaires » peuvent apparaître pour d'autres raisons que l'absorption de rayonnement lumineux. Par exemple, une jonction semi-conductrice *p - n*, polarisée à l'aide d'une source de tension continue, peut « injecter » dans le semi-conducteur des paires électron-trou supplémentaires : dans ce cas, la luminescence qui en résulte est une conversion d'énergie électrique en énergie lumineuse : c'est un phénomène d'**électroluminescence.**

Les rayonnements de grande énergie, rayons γ, particules diverses de la physique nucléaire, peuvent créer dans certains semi-conducteurs des paires électron-trou; en recueillant la lumière émise par luminescence, on a un moyen de détecter la présence de rayonnement de grande énergie, et même de compter des particules et de mesurer leurs énergies, d'où l'emploi de certains cristaux dits « scintillateurs » en physique nucléaire.

Les phénomènes varient d'un semi-conducteur à l'autre, dépendent des impuretés et des défauts, se modifient avec la température, les conditions d'excitation, etc. Il y a là une grande diversité de situations expérimentales et, par suite, une quantité énorme d'expériences à effectuer pour rendre compte des phénomènes observés, en même temps qu'une grande richesse d'applications possibles.



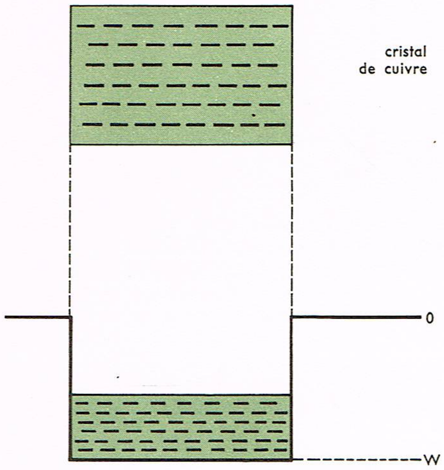
La recombinaison d'un électron d'énergie E_2 et d'un trou d'énergie E_1 produit l'émission d'un photon d'énergie $h\nu = E_2 - E_1$.



La capture d'un électron d'énergie E_2 sur un niveau d'impureté d'énergie E_i peut s'accompagner de l'émission d'un photon d'énergie $h\nu = E_2 - E_i$; dans un deuxième temps, un photon d'énergie $h\nu = E_i - E_1$ peut être émis.

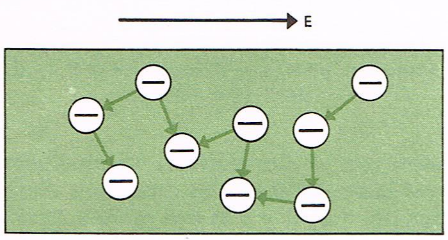
Les semi-conducteurs ont bien d'autres propriétés, toutes aussi passionnantes que celles qui viennent d'être sommairement décrites, et dont nous n'avons pas la possibilité de parler : propriétés magnétiques, thermiques, effets thermoélectriques, etc.

Tous ces phénomènes sont, en général, en bon accord avec la théorie. Si quelques obscurités subsistent, c'est parce que la complexité des choses dépasse encore souvent nos moyens d'expérience ou nos possibilités de calcul, et non pas en raison de quelque mystère caché ou de quelque difficulté fondamentale.



Modèle simplifié d'un métal.

En prenant pour origine l'énergie d'un électron au repos en dehors du métal, on peut admettre que les électrons de ce métal sont dans un puits de potentiel de profondeur W. A température assez élevée, quelques électrons ont une énergie cinétique supérieure au puits de potentiel et peuvent s'échapper du métal : c'est l'émission thermo-électronique.



Mouvement d'un électron dans un métal.

Un électron dans un métal, soumis à un champ électrique E, se déplace en moyenne en sens inverse du champ. Ce déplacement est la résultante d'une série de portions de trajectoires de longueurs variables, interrompues par des chocs survenant au hasard; le temps moyen τ qui sépare deux chocs est appelé le *temps de relaxation*. Faisons un calcul numérique dans le système M. K. S. A. (Mètre, Kilogramme, Seconde, Ampère); dans la formule

$$\sigma = \frac{ne^2\tau}{m},$$

la conductivité électrique du cuivre σ est $6 \cdot 10^7$ (ohm-m)$^{-1}$, le nombre d'électrons par mètre cube est à peu près 10^{29}, la masse de l'électron m vaut 10^{-30} kilogramme et sa charge $e = 1{,}6 \cdot 10^{-19}$ coulomb; on en déduit

$$\tau = \frac{6 \cdot 10^7 \times 10^{-30}}{10^{29}(1{,}6 \cdot 10^{-19})^2} = 2 \cdot 10^{-14}\ \text{s}.$$

Comme la vitesse moyenne u est, à température ordinaire, de 10^6 m/s, la distance moyenne entre deux chocs, ou libre parcours moyen λ, est d'environ
$\lambda = uT = 10^6 \times 2 \cdot 10^{-14} = 2 \cdot 10^{-8}\ m = 200$ Å.

Un métal est un solide dont une des bandes d'énergie, la bande de conduction, est partiellement remplie d'électrons. Par exemple, la bande de conduction d'un métal monovalent, comme le cuivre, possède 2N places (N est le nombre d'atomes du cristal considéré) où s'entassent N électrons. L'existence même des bandes d'énergie est une conséquence de la mécanique quantique appliquée aux cristaux.

de l'étude de la conductivité électrique des métaux à l'explication inespérée du libre parcours des électrons

Comment se représentait-on un métal avant la théorie des bandes? Depuis longtemps, on savait par expérience que le cuivre, par exemple, était un excellent conducteur de l'électricité; sa conductivité, trop grande pour être d'origine ionique, devait être attribuée au mouvement des électrons; on supposait donc l'existence d'une sorte de gaz d'électrons libres à l'intérieur du métal et contenus dans un puits de potentiel de profondeur — W.

Ce modèle très simple rend compte de certaines propriétés des métaux. Efforçons-nous, par exemple, d'expliquer la conductivité électrique du cuivre. Sous l'influence d'un champ électrique E constant, un électron de masse m, de charge $-e$ prend une accélération $\vec{\gamma} = \frac{d\vec{v}}{dt}$ donnée par la formule

$$m\frac{d\vec{v}}{dt} = -e\vec{E}.$$

Par intégration, on en déduit

$$\vec{v}(t) = \vec{v}(0) - (e/m)\,\mathrm{E}t,$$

où $\vec{v}(t)$ est la vitesse de l'électron à l'instant t, et $\vec{v}(0)$ sa vitesse à l'instant initial ($t = 0$). Un tel mouvement, caractérisé par une vitesse croissant indéfiniment avec le temps, ne peut se prolonger très longtemps : l'électron dont nous étudions le mouvement n'est pas seul dans la boîte! Il y a avec lui beaucoup d'autres électrons, et surtout des ions cuivre; on peut penser qu'au bout d'un temps τ, appelé « temps de relaxation », l'électron subit un choc qui perturbe complètement son état de mouvement; cet électron, au temps τ, juste avant le choc, transporte un courant élémentaire égal à

$$-e\vec{v}(\tau) = -e\vec{v}(0) + (e^2/m)\vec{E}\tau;$$

la densité de courant $\vec{j}$, ou courant électrique traversant une surface de section unité, s'obtient en ajoutant les courants élémentaires produits par les électrons situés dans un volume unité, d'où

$$\vec{j} = -e\,\Sigma\,\vec{v}(\tau) = -e\,\Sigma\,\vec{v}(0) + \Sigma\,(e^2/m)\,\vec{E}\tau,$$

où le signe Σ indique que l'on ajoute tous les vecteurs vitesse $\vec{v}(\tau)$. Si l'on suppose que la vitesse de l'électron après choc est quelconque, c'est-à-dire déterminée par le hasard sans aucune direction privilégiée, il en est de même pour la vitesse initiale $\vec{v}(0)$, et la somme $\Sigma\,\vec{v}(0)$ est nulle puisque c'est une somme de vecteurs pointant au hasard dans toutes les directions possibles.

Si l'on désigne par n le nombre d'électrons par unité de volume, la somme $\Sigma\, e^2/m\,\vec{E}\tau$ s'écrit $n\ e^2/m\ \vec{E}\tau$, et l'on a

$$\vec{j} = \frac{ne^2\tau}{m}\vec{E}.$$

Cette relation montre que la densité de courant $\vec{j}$ est proportionnelle au champ électrique $\vec{E}$; on retrouve la célèbre loi expérimentale d'Ohm (1827),

$$\vec{j} = \sigma\,\vec{E},$$

où σ est la conductivité électrique; en identifiant les deux expressions, on en déduit $\sigma = \frac{ne^2\tau}{m}$.

Nous verrons plus loin que la vitesse moyenne u d'un électron dans le cuivre est d'environ 10^6 mètres par seconde, et que le temps de relaxation est d'environ 2×10^{-12} seconde, de sorte qu'un électron franchit entre deux chocs une distance (libre parcours moyen) égale à : $\lambda = u\tau = 10^6\ \text{m/s} \times 2 \cdot 10^{-14}\ \text{s} = 2 \cdot 10^{-8}$ m $= 200$ angströms.

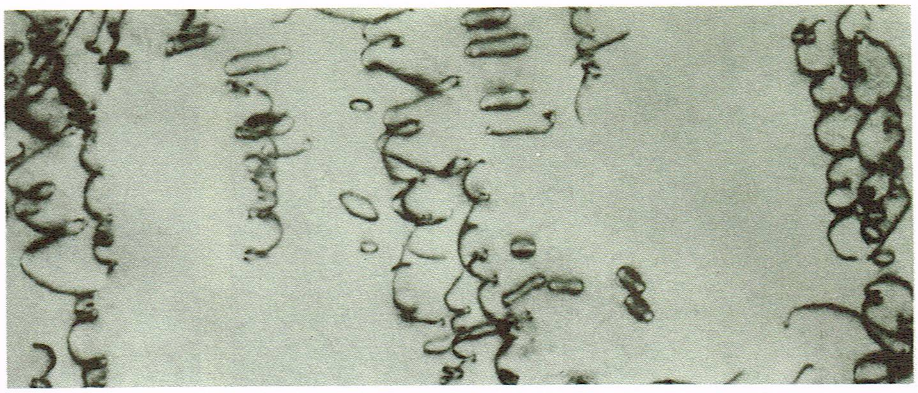
Or, la distance entre deux atomes de cuivre est de l'ordre de quelques angströms, ce qui signifie qu'**un électron franchit une centaine de distances interatomiques avant de subir un choc.**

D'autres expériences, que nous ne pouvons décrire ici, permettent d'affirmer que cette évaluation du libre parcours

moyen est correcte. Et pourtant, dans un métal comme le cuivre, les ions Cu^+ sont empilés les uns contre les autres, ne laissant entre eux que de petits interstices, et il est difficilement concevable qu'un électron puisse franchir en ligne droite une distance d'une centaine de distances interatomiques! **La mécanique ondulatoire permet d'expliquer ce paradoxe.** L'électron n'est pas un corpuscule étroitement localisé, mais est piloté par une onde de probabilité; cette onde, caractérisée par un vecteur d'onde $\vec{k}$ et par une pulsation ω, rencontre les atomes du réseau cristallin sur lesquels elle se diffracte; si $\vec{k}$ et ω sont quelconques, la diffraction de l'onde en tous sens rend la propagation impossible; mais, de même que la propagation des rayons X ne se produit que si une certaine condition, la condition de Bragg, est satisfaite, de même l'onde électronique ayant un vecteur d'onde donnée $\vec{k}$, ne se propagera que pour une valeur de la pulsation ω, dépendant de $\vec{k}$. Nous avons appris précédemment que la pulsation ω de l'onde pilote d'une particule est liée à l'énergie E de cette particule par la relation $E = h\omega$, où h est la constante de Planck. On voit que l'énergie E est une fonction du vecteur d'onde $\vec{k}$; nous retrouvons la notion d'énergie permise et d'énergie interdite.

Si la périodicité du cristal était rigoureusement parfaite, l'onde pilote d'énergie $E = h\omega$ adéquate se propagerait indéfiniment et se réfléchirait simplement sur les surfaces extérieures du cristal; en d'autres termes, le libre parcours moyen λ, au lieu d'être de l'ordre de grandeur de quelques centaines d'angströms, ne serait limité que par les dimensions des cristaux; la conductivité électrique serait énorme. En fait, la présence de défauts et le mouvement thermique des atomes du réseau en brisent l'exacte périodicité, et l'onde électronique a un libre parcours limité. C'est ce que confirme très précisément l'expérience : la conductivité à température fixe s'accroît si le cristal contient moins d'impuretés, elle augmente lorsque en refroidissant le cristal donné on diminue l'amplitude des vibrations thermiques du réseau cristallin.

En somme, la théorie simpliste des électrons libres dans les métaux donne un résultat honorable et assez inespéré pour le calcul de la conductivité électrique des métaux.



Ces curieux graffiti montrent en réalité une dislocation en hélice dans un alliage aluminium-cuivre.

Mais la comparaison avec l'expérience donne pour le libre parcours moyen des valeurs que cette théorie est incapable d'expliquer. La mécanique quantique appliquée au solide nous montre pourquoi une onde électronique peut franchir une aussi grande distance dans un espace rempli d'ions et nous fait mieux comprendre la nature des interactions qui conduisent aux chocs des électrons.

quelques propriétés des métaux ou la querelle des statistiques

Dans le paragraphe précédent, nous n'avons pas fait intervenir la notion de statistique ou de probabilité d'occupation (ce que nous aurions dû faire en toute rigueur), pour ne pas alourdir l'argument et bien dégager les idées essentielles. Pour certaines questions, au contraire (chaleur spécifique des électrons, pouvoir thermo-électronique des métaux, etc.), la statistique joue un rôle capital.

Les mouvements spontanés des électrons confèrent au gaz électronique une certaine énergie ν_e d'autant plus grande que les mouvements sont plus rapides, c'est-à-dire que la température est plus élevée. On sait que la chaleur spécifique à volume constant d'un système est la quantité de chaleur, ou énergie calorifique, qu'il faut fournir au système pour en élever la température d'un degré; la chaleur spécifique des électrons est donc le rapport $\delta\nu_e$ de l'énergie du gaz électronique à l'accroissement de température δT qui lui correspond :

$$Cv = \frac{\delta\nu e}{\delta T} = \frac{\delta\nu}{\delta T}.$$

Pour calculer C_v, il faut évaluer l'énergie ν_e correspondant au mouvement thermique des électrons; il faut pour cela connaître la distribution statistique des vitesses des électrons ou, comme on dit, « faire la statistique du gaz d'électrons ».

Le pouvoir thermo-électronique d'un métal est son aptitude à émettre spontanément des électrons lorsqu'il est porté dans le vide à une température assez élevée.

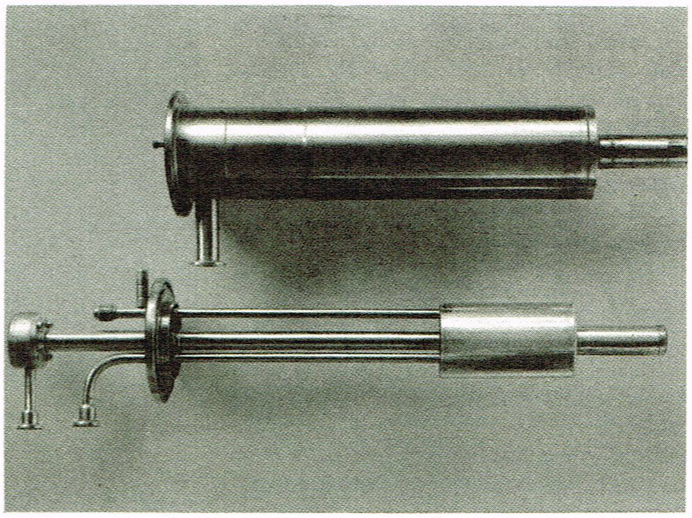
Dans le modèle simple des électrons dans un puits de potentiel, certains électrons, ayant acquis par agitation thermique une énergie positive, peuvent s'échapper du puits et sortir du métal. Là encore nous devons résoudre un problème de statistique. L'expérience montre que le *courant émis* par émission thermo-électronique est de la forme $I = SA \exp(-E/kT)$, où S est l'aire de la surface émettrice, T sa température absolue, A une constante variant faiblement avec la température, et la constante E une « énergie d'activation » de l'ordre de quelques électron-volts.

La statistique de Maxwell-Boltzmann déjà cité, imaginée vers la fin du XIXe siècle à propos de la théorie cinétique des gaz, a été appliquée au modèle des électrons libres dans un métal. En particulier, cette statistique permet de calculer la chaleur spécifique électronique d'un métal ou son pouvoir thermo-électronique. On obtient une chaleur spécifique très supérieure à la valeur expérimentale. Pour le pouvoir thermo-électronique, l'accord avec l'expérience n'est pas meilleur : l'énergie d'activation mesurée, E, est différente de celle que la théorie permet de prévoir.

En fait, la mécanique quantique implique l'indiscernabilité des particules et entraîne

Doc. C. N. R. S.





La physique des basses températures.

L'homme sait, depuis le début du siècle, obtenir en laboratoire des « froids » beaucoup plus rigoureux que ceux que la nature présente dans certaines contrées. Pour atteindre les très basses températures plutôt que d'affronter l'ours blanc, mieux vaut disposer d'un cryostat à hélium liquide. A la température de liquéfaction de ce gaz (4 degrés absolus, c'est-à-dire 269 degrés Celsius au-dessous de zéro), les solides présentent des propriétés nouvelles, dont la plus remarquable, la *supraconductivité*, est longtemps restée mystérieuse.

le rejet de la statistique de Maxwell-Boltzmann, fondée précisément sur le caractère discernable des particules; cette statistique doit être remplacée soit par celle de Bose-Einstein, soit par celle de Fermi-Dirac. La statistique de Bose-Einstein, nous l'avons vu par ailleurs, appliquée aux photons du corps noir, permet de prévoir avec exactitude son pouvoir émissif; et nous avons vu que cette statistique appliquée aux phonons, ou quanta de vibration des réseaux cristallins, permet d'expliquer correctement leur chaleur spécifique à basse température. Pour les particules qui, comme les électrons, obéissent au principe de Pauli, c'est-à-dire ne peuvent être plusieurs dans le même état quantique, la statistique à utiliser est celle de Fermi-Dirac. Le pouvoir thermo-électronique s'explique alors de façon très satisfaisante; c'est d'ailleurs cette statistique que nous avons appliquée au problème général de l'occupation des bandes d'un solide par les électrons.

une énigme longue à dissiper : la supraconductivité

Trois ans seulement après avoir liquéfié l'hélium pour la première fois, Kamerlingh Onnes découvre en 1911 la supraconductivité : la résistivité électrique du mercure devient subitement nulle au-dessous de 4,2 degrés absolus (c'est-à-dire 268,8 °C au-dessous de zéro); le passage de l'état normal à l'état supraconducteur se produit brutalement en quelques centièmes de degré!

Nous avons vu qu'à basse température le libre parcours moyen d'un électron devient grand et que la conductivité électrique d'un métal s'accroît, ou, si l'on préfère, que sa résistivité électrique décroît. Néanmoins, en extrapolant au zéro absolu on constate que la résistivité doit tendre vers une valeur résiduelle : d'inévitables défauts existent toujours dans un cristal, qui doivent empêcher la résistivité électrique de s'annuler au zéro absolu; avant la découverte de Kamerlingh Onnes, c'est ce que l'expérience avait toujours montré. L'exception allait-elle confirmer la règle? Nullement! Bientôt l'extraordinaire propriété du mercure cessait d'être exceptionnelle : aux très basses températures, le plomb, l'étain, l'indium, se révélaient supraconducteurs! Aujourd'hui on connaît vingt-deux éléments et près de trois cents composés supraconducteurs.

Le phénomène de la supraconductivité, connu maintenant depuis plus d'un demi-siècle, est encore dans une certaine mesure incompréhensible. Depuis cinquante ans, des physiciens toujours plus nombreux étudient la supraconductivité et, plus généralement, les propriétés associées à l'état supraconducteur. Jusqu'en 1950, seuls les travaux expérimentaux avaient pu apporter quelques satisfactions aux chercheurs, et aucune interprétation théorique valable n'avait été proposée. Nous allons d'abord décrire brièvement les faits expérimentaux les plus marquants de l'état supraconducteur, avant d'indiquer dans quelle direction la théorie a réalisé récemment quelques progrès, importants et peut-être même décisifs.

Supposons qu'un anneau d'un métal supraconducteur plongé dans un champ magnétique soit refroidi au-dessous de sa température de transition de façon à être amené dans l'état supraconducteur; la suppression du champ magnétique lance dans l'anneau un courant d'induction qui doit persister indéfiniment si la résistance de l'anneau est nulle; un courant de plusieurs centaines d'ampères a circulé dans un anneau supraconducteur d'un laboratoire du Massachusetts Institute of Technology pendant plusieurs années sans manifester la moindre décroissance!

On s'aperçut assez rapidement qu'un champ magnétique suffisamment intense peut supprimer la supraconductivité en restituant au métal une résistivité non nulle, en le ramenant à l'état normal; on appelle « champ magnétique critique H_c (T) », la valeur du champ magnétique qui, à la température T, supprime l'état supraconducteur.

Il en résulte une conséquence importante : la supraconductivité d'un fil métallique disparaît d'elle-même lorsque circule un courant trop important (lorsque la densité de courant dans le fil y crée un champ magnétique égal au champ critique).

Phot. Klages, Atlas-Photo et C. N. E. T.

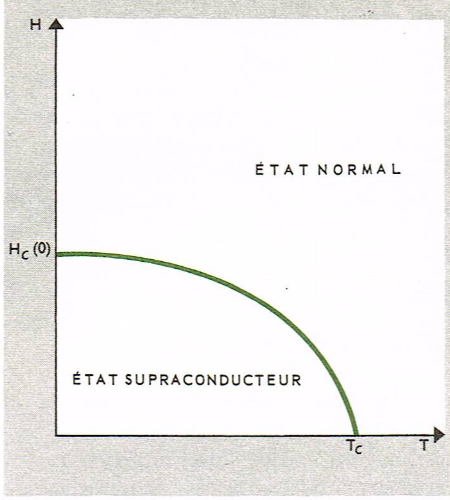
On constate que, pour différents supraconducteurs, la température critique T_c est d'autant plus élevée que le champ critique au zéro absolu de température $H_c(0)$ est plus grand. La plupart des supraconducteurs ont des températures critiques T_c qui varient de quelques degrés absolus à une petite fraction de degré; exceptionnellement, certains alliages ont une température critique relativement élevée; le record (18 °K) est détenu par le niobiure d'étain Nb^3Sn.

L'effet isotopique découvert en 1950 est d'une importance particulière : la mesure précise de la température critique des différents isotopes du mercure a montré que cette température dépendait légèrement de la masse atomique; on a trouvé la loi empirique suivante : $(M)^{1/2}T_c =$ constante.

Or, la fréquence des phonons, ou quanta de vibration des réseaux, est inversement proportionnelle à $(M)^{1/2}$; cela suggère évidemment que les phonons jouent un rôle important dans l'état supraconducteur, ce sur quoi nous reviendrons.

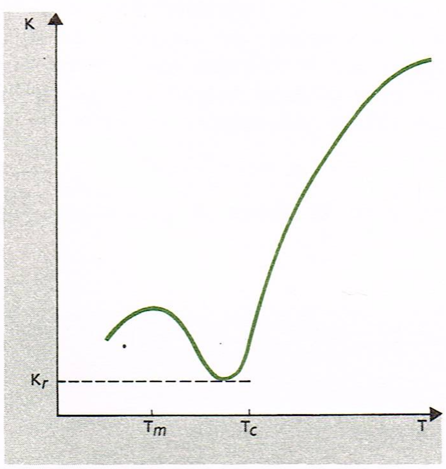
Un métal sous la forme d'un monocristal très pur, sans contraintes élastiques, passe de l'état normal à l'état supraconducteur dans une gamme très faible de températures, quelques centièmes de degré Kelvin : la température critique est bien définie; l'expérience montre que cette température varie avec la teneur en impuretés et l'état de contrainte élastique. On constate aussi que, pour un matériau *non homogène*, le domaine de transition est beaucoup plus large : les différentes régions de l'échantillon ayant des températures de transition différentes, le passage de l'état supraconducteur à l'état normal s'établit progressivement.

Même pour les supraconducteurs très homogènes, il a été prouvé que, lors du passage de l'état supraconducteur à l'état normal, certaines parties du métal sont encore dans l'état supraconducteur, tandis que les autres ont repris leur état normal. Cela a permis de préciser l'idée déjà avancée vers 1935, suivant laquelle les électrons se partageraient en deux races, les électrons supraconducteurs et les électrons normaux, la proportion des uns et des autres changeant avec la température qui règne dans la zone où se fait la transition d'un état à l'autre. L'existence de régions, prouvée expérimentalement, montre d'ailleurs que les électrons supraconducteurs ont tendance à se grouper en domaines d'où les électrons normaux sont exclus.



A une température T et en présence d'un champ magnétique $\vec{H}$, observons si un métal est supraconducteur ou normal. Si l'on traduit le résultat de ces observations sur un diagramme température-champ magnétique, on voit que les régions correspondant à l'état supraconducteur sont séparées des régions correspondant à l'état normal par une courbe $H_c(T)$.

L'étude de la conductivité thermique des métaux supraconducteurs a montré que cette conductivité s'abaisse fortement quand on est au-dessous de la température critique; les électrons supraconducteurs sont donc incapables de conduire la chaleur; la conductivité thermique résiduelle K_r est due au transport de chaleur par les phonons du réseau, et la croissance de cette conductivité au voisinage de T_m montre que ce transport de chaleur par les phonons devient de plus en plus facile : les phonons, dont le mouvement est normalement gêné par la présence d'électrons, ignorent donc les électrons supraconducteurs; on conçoit donc qu'à charge de revanche les électrons supraconducteurs dédaignent à leur tour les phonons et que, par suite, rien ne s'oppose au passage de l'électricité. Remarquons d'ailleurs qu'au-dessous de T_m la conductivité thermique retombe à zéro, car il n'y a plus de phonons excités.



La conductivité thermique des métaux supraconducteurs.

Le diagramme donnant la conductivité thermique K d'un métal supraconducteur en fonction de la température au voisinage de la transition montre que les électrons supraconducteurs ne conduisent pas la chaleur.

Dans le dédale des propriétés de l'état supraconducteur, un peu d'ordre commence donc à apparaître : existence de domaines; possibilité pour les électrons d'être dans deux états : l'état supraconducteur ou l'état normal. Les électrons supraconducteurs n'interagissent pas avec le réseau cristallin, bien que l'état même de supraconductivité soit (comme le suggère l'effet isotopique) lié à l'interaction électron-réseau.

Depuis 1950, l'étude de la supraconductivité par la mécanique quantique s'est beaucoup développée; en particulier, la théorie de Bardeen, Cooper, Schriefer a réalisé un progrès majeur : suivant eux, deux électrons peuvent être couplés par l'intermédiaire du réseau cristallin pour former une paire; tous les électrons supraconducteurs sont associés pour former de telles paires. Cette théorie permet d'expliquer certains des faits expérimentaux que nous avons examinés plus haut, et prévoit aussi l'existence d'une bande interdite de largeur très faible de l'ordre de kT_c (c'est-à-dire 10^{-3} eV pour $T_c = 6$ °K); l'expérience a confirmé la théorie. Pour la première fois dans l'histoire de la supraconductivité, la théorie reprenait un léger avantage en faisant une prédiction que l'expérience a confirmée, alors que, pendant cinquante ans, l'expérience avait accumulé des faits que les théoriciens ne pouvaient expliquer.

propriétés magnétiques des solides

L'importance théorique du magnétisme est considérable : le lecteur sait que de nombreuses particules élémentaires possèdent en propre un moment magnétique : le spin; cela signifie que le magnétisme est un attribut premier de la nature, au même titre que la charge électrique élémentaire.

D'ailleurs, le simple mouvement de ces charges électriques crée aussi un champ magnétique. De là résultent pour les solides d'innombrables propriétés magnétiques dont la technique ne manque pas de tirer parti. On va se borner à décrire quelques-unes de ces propriétés et à indiquer quelques applications qui en découlent.

Connaître les propriétés magnétiques d'un solide, c'est, dans une large mesure, connaître l'aimantation $\vec{J}$ que prend ce solide lorsqu'il est soumis à un champ magnétique d'origine extérieure $\vec{H}$; on pose en général que $\vec{J}$ est proportionnel à $\vec{H}$: $\vec{J} = \chi\vec{H}$; c'est le coefficient de proportionnalité χ, appelé « susceptibilité magnétique », qu'il s'agit de calculer.

Certains solides présentent une susceptibilité χ négative, très faible, indépendante de la température : on dit qu'ils sont « diamagnétiques »; ce comportement est dû aux électrons qui modifient leur mouvement de façon à s'opposer, partiellement, au flux créé par un champ magnétique extérieur; c'est en somme la loi d'induction, ou loi de Lenz, à l'échelle des particules.

Les autres solides présentent une susceptibilité positive, faible et dépendant fortement de la température : on dit qu'ils sont « paramagnétiques ». Il existe une exception importante : certains solides montrent un paramagnétisme géant; on les appelle « ferromagnétiques ».

magnétisme nucléaire

Les noyaux, formés de protons et de neutrons, particules élémentaires possédant un spin, ont eux-mêmes un moment magnétique dit « spin nucléaire ». Comme ces moments magnétiques nucléaires sont environ mille fois plus faibles que le moment magnétique dû au spin de l'électron, leur contribution à la susceptibilité magnétique χ est en général tout à fait négligeable; ils jouent pourtant un rôle important dans le phénomène sélectif de la résonance magnétique, que nous examinerons plus loin. Pour l'instant nous ne considérerons que le magnétisme électronique.

le théorème de Larmor et son application au magnétisme

En utilisant le modèle, commode, d'atome simplifié de Bohr, il est facile de voir comment un champ magnétique agit sur les électrons d'un atome; la fréquence propre de la trajectoire d'un électron sur son orbite est modifiée de telle façon que le nouveau mouvement crée un petit moment magnétique μ_e opposé au champ magnétique extérieur :

$$\mu_e = -\frac{e^2H}{4mc^2}r^2,$$

où e est la charge de l'électron, m sa masse, c la vitesse de la lumière, H le champ magnétique extérieur, et r le rayon de l'orbite de l'électron. Le moment magnétique μ_{at} dû aux Z électrons de l'atome envisagé s'écrit

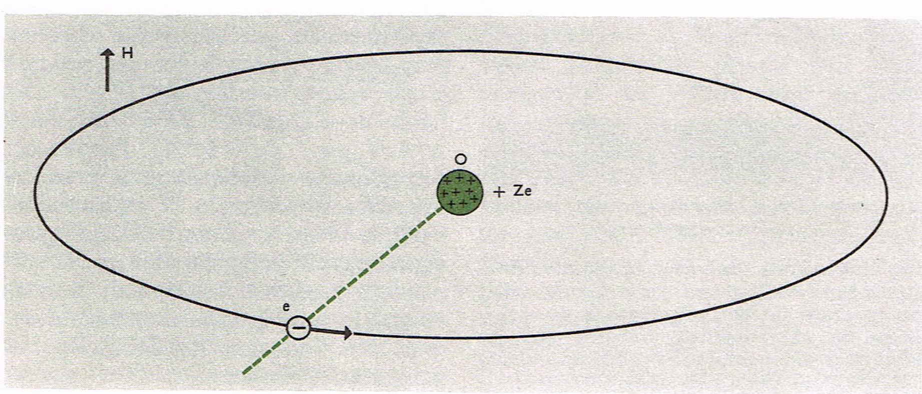
$$\mu_{at} = -\frac{Ze^2H}{4mc^2} < r^2 >,$$

où $< r^2 >$ est la moyenne du carré de r, moyenne prise sur les différentes orbites électroniques et en tenant compte de ce que toutes les orbites ne sont pas perpendiculaires au champ magnétique. Si l'on désigne par N le nombre d'atomes par unité de volume, l'aimantation $\vec{J}$ prend la valeur $\vec{J} = -\frac{Ze^2N}{4mc^2} < r^2 > \vec{H}$, et, par suite, la susceptibilité diamagnétique s'écrit

$$\chi_{\text{dia}} = -\frac{Ze^2N}{4mc^2} < r^2 >.$$

On trouve par cette formule des valeurs de χ négatives valant de 10^{-6} à 10^{-5} et conformes à l'expérience. (Dans le calcul fait ci-dessus, on a considéré un solide comme la simple juxtaposition d'atomes, puisqu'on a ajouté la contribution de chacun d'entre eux.)

Dans un cristal, cette méthode est tout à fait justifiée pour les électrons des bandes profondes, qui correspondent à des orbites atomiques à peine modifiées; au contraire, pour les électrons de valence d'un cristal, on ne peut plus parler d'orbites autour d'un atome donné, et la méthode ci-dessus n'est plus valable. Le problème doit être traité par les méthodes plus élaborées de la mécanique quantique. De toute façon, le diamagnétisme, quelle que soit la méthode utilisée pour l'évaluer, apparaît toujours comme un phénomène de



Dans le modèle d'atome simplifié dû à Bohr, un électron (e) de masse m décrit autour du noyau (O) un mouvement circulaire de pulsation ω_0. La présence d'un champ magnétique $\vec{H}$ modifie le mouvement de l'électron; on montre que la nouvelle pulsation ω est égale à

$$\omega = \omega_0 + \omega_M,$$

avec $\omega_M = -\frac{eH}{2mc}$ (c étant la vitesse de la lumière), et que cette modification du mouvement est équivalente à l'apparition d'un moment magnétique induit :

$$\mu = -\frac{e^2Hr^2}{4mc^2},$$

où r est le rayon de l'orbite.



Diamagnétisme.

En l'absence de champ magnétique, tous les moments magnétiques deux à deux opposés dans chaque atome (pour simplifier on a représenté un seul atome) ont une résultante évidemment nulle. En présence d'un champ magnétique $\vec{H}$, tous les moments subissent la même modification parallèle au champ magnétique et de sens contraire (sur la partie droite de la figure on a représenté en trait vert les deux moments magnétiques d'un atome avant l'apparition du champ, et en trait noir les deux moments en présence du champ); la résultante générale, l'aimantation $\vec{J}$, est donc parallèle et opposée au champ magnétique.

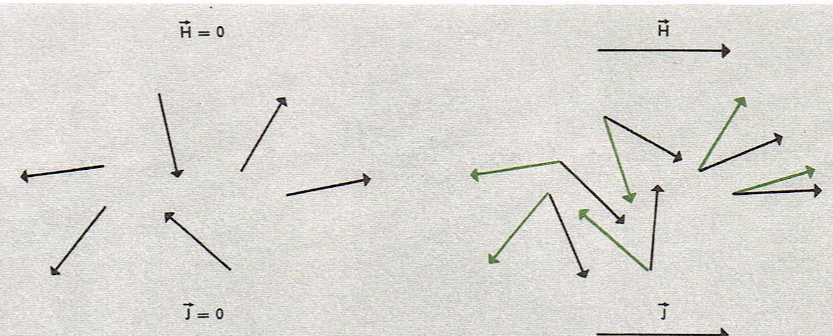
réaction : les électrons tendent à modifier leurs mouvements de façon à s'opposer au flux créé par le champ magnétique extérieur; on retrouve un aspect microscopique de la loi d'induction, ou loi de Lenz.

le magnétisme d'orientation

Il existe deux types de moments magnétiques permanents : ceux qui sont liés aux spins des particules élémentaires, et ceux qui sont dus à des mouvements électroniques.

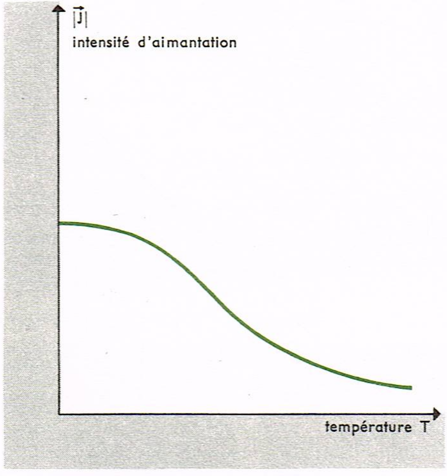
Les premiers, si l'on excepte les spins nucléaires dont nous avons déjà parlé, sont dus aux spins des électrons; le nombre quantique de spin électronique ne peut prendre que deux valeurs, $\pm 1/2$, de sorte que le moment magnétique propre de l'électron suivant un axe quelconque ne peut prendre que les deux valeurs $\pm \mu_B$ ($\mu_B = \frac{eh}{2mc}$ est appelé le « magnéton de Bohr).

Les moments magnétiques du second type, dans le cas d'un atome, sont dus aux mouvements orbitaux des électrons et sont liés aux moments cinétiques correspondants. Pour les électrons des bandes profondes d'un cristal, cette représentation, nous l'avons dit, conserve un sens; mais leurs bandes d'énergie sont pleines et on démontre que le moment magnétique résultant est nul : le magnétisme d'orientation des couches profondes disparaît. Pour les électrons d'une bande non pleine d'énergie élevée on ne peut plus parler d'orbites et l'on doit avoir recours aux méthodes plus puissantes de la mécanique quantique. On cherche comment un champ magnétique modifie l'état dynamique des électrons d'un cristal; on trouve en général deux contributions : l'une, négative, correspond au diamagnétisme; l'autre, positive, traduit le paramagnétisme.



Magnétisme d'orientation.

En l'absence de champ, tous les moments magnétiques atomiques supposés identiques (pour raison de simplicité on en a représenté seulement six) sont orientés au hasard : leur résultante, l'aimantation, est nulle. En présence d'un champ magnétique $\vec{H}$, ces moments tendent à devenir parallèles au champ (on a représenté, sur la partie droite de la figure, en trait vert les positions des moments qui précèdent l'apparition du champ, et en trait noir les positions après aimantation sous l'influence du champ); la résultante, l'aimantation, est donc parallèle au champ et de même sens.



Courbe d'aimantation d'une substance paramagnétique en fonction de la température pour une intensité donnée du champ magnétique.

La statistique de Maxwell-Boltzmann montre que

$$N_2/N_1 = \exp - \mu H/kT.$$

A partir de cette relation, et en tenant compte de ce que $N_1 + N_2 = N$, on peut calculer N_1 et N_2 :

$$N_1 = N \frac{\exp \vec{\mu}\vec{H}/kT}{\exp \vec{\mu}\vec{H}/kT + \exp - \vec{\mu}\vec{H}/kT}$$

$$N_2 = N \frac{\exp - \vec{\mu}\vec{H}/kT}{\exp \vec{\mu}\vec{H}/kT + \exp - \vec{\mu}\vec{H}/kT}.$$

L'aimantation $\vec{J}$ s'écrit

$$\vec{J} = N\vec{\mu} \frac{\exp \vec{\mu}\vec{H}/kT - \exp - \vec{\mu}\vec{H}/kT}{\exp \vec{\mu}\vec{H}/kT + \exp - \vec{\mu}\vec{H}/kT}.$$

A basse température, $\vec{\mu}\vec{H}/kT$ est beaucoup plus grand que l'unité, et l'on a $\vec{J} = N\vec{\mu}$; l'aimantation est maximale parce que les moments magnétiques élémentaires sont tous orientés dans le sens du champ magnétique.

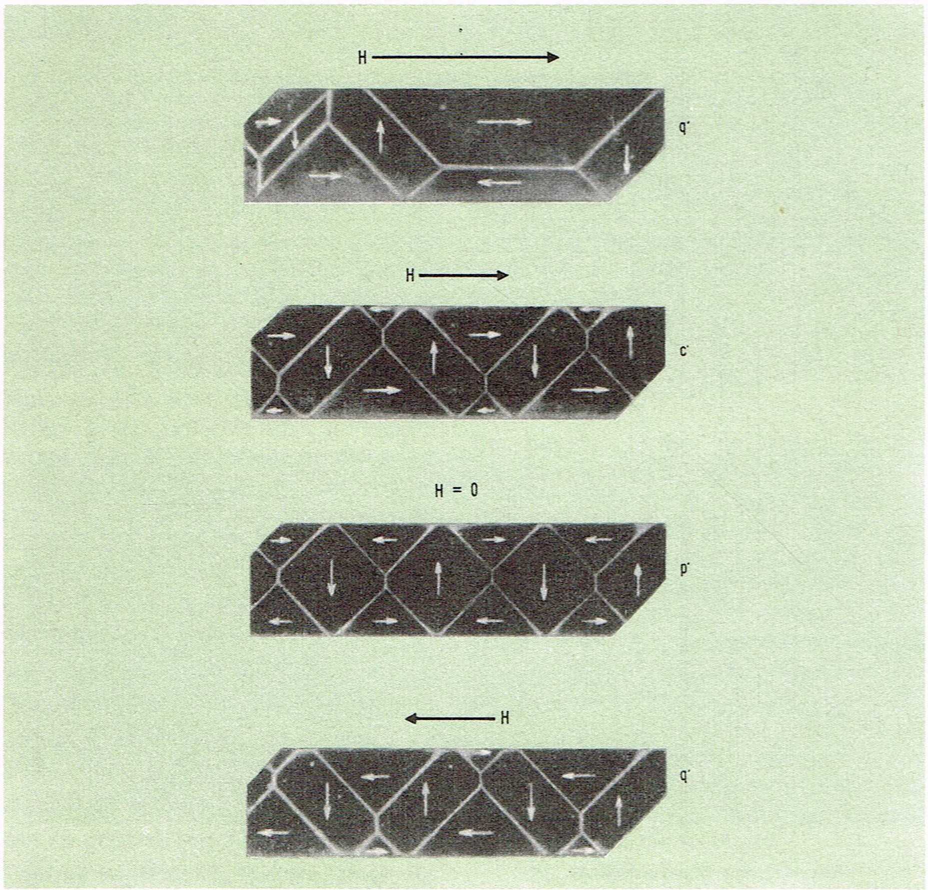
A température élevée, $\vec{\mu}\vec{H}/kT$ est plus petit que l'unité, et l'on a

$$\vec{J} = N\vec{\mu}^2/kT\,\vec{H}.$$

La susceptibilité paramagnétique

$$\chi_{par} = N\vec{\mu}^2/kT$$

est positive et décroît avec la température. C'est ce que la courbe expérimentale ci-contre confirme.



Mouvement des parois de domaines ferromagnétiques.

Tous les moments magnétiques dus soit à des spins électroniques, soit à des mouvements électroniques, tendent à s'orienter parallèlement au champ magnétique extérieur et à donner ainsi une contribution positive à la susceptibilité magnétique : c'est le paramagnétisme; lorsque, pour un solide, tous les moments permanents se compensent exactement, il subsiste le diamagnétisme sous-jacent dont nous avons parlé. Revenons au cas des solides paramagnétiques. De nombreuses causes peuvent favoriser ou contrarier l'orientation des moments magnétiques : leurs interactions mutuelles, les chocs avec les électrons du solide et avec les phonons du réseau, le couplage avec les impuretés, etc. On conçoit que le problème du magnétisme d'orientation soit très difficile à résoudre.

Un modèle simple permet pourtant de traiter très approximativement le paramagnétisme. Nous admettrons que, dans un système contenant N moments magnétiques par unité de volume, chaque moment élémentaire peut s'orienter parallèlement à un champ magnétique extérieur $\vec{H}$, soit dans le sens du champ, soit en sens contraire. Supposons que l'on connaisse le nombre N_1 de moments qui s'orientent dans le sens du champ et le nombre N_2 de moments qui s'orientent en sens contraire (avec $N_1 + N_2 = N$).

L'aimantation $\vec{J}$, résultante de ces moments, est donnée par :

$$\vec{J} = N_1 \vec{\mu} - N_2 \vec{\mu}.$$

Un calcul simple de statistique permet précisément d'obtenir N_1 et N_2 à toute température T. Le résultat ainsi trouvé est en bon accord avec les courbes expérimentales donnant l'aimantation paramagnétique en fonction de la température.

le ferromagnétisme

Certains solides ont un comportement magnétique anormal : ils sont doués d'un « paramagnétisme géant », avec des valeurs de la susceptibilité magnétique χ de l'ordre de 100, 1 000, ou davantage. Cette propriété est appelée « ferromagnétisme », parce que le fer est l'un des matériaux les plus usuels qui la présente.

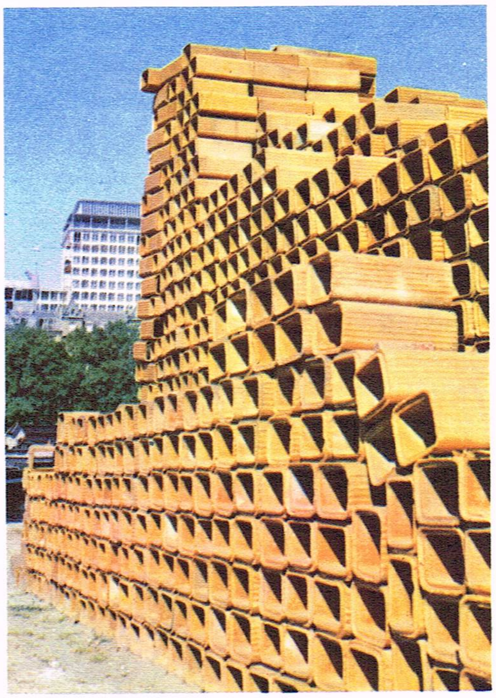
En 1907, Pierre Weiss, pour expliquer le comportement des corps ferromagnétiques, avance l'hypothèse que, dans un cristal ferromagnétique, tous les spins électroniques d'une région du cristal, par suite de l'action d'un certain champ cristallin, sont systématiquement parallèles entre eux : une telle région, appelée « domaine », doit donc posséder un moment magnétique énorme. A l'état de repos, sans champ magnétique, l'aimantation résiduelle peut être assez faible, parce que les moments magnétiques des différents domaines, orientés en tous sens, se compensent. Sous l'influence d'un champ magnétique extérieur, le moment magnétique de chaque domaine tend à basculer pour prendre la même direction que ce champ magnétique extérieur. Le comportement « paramagnétique géant » en découle évidemment, puisque le mécanisme est semblable à celui du paramagnétisme, mais appliqué à des moments magnétiques beaucoup plus grands. Quelques années plus tard, cette hypothèse fut confirmée par l'expérience; les domaines, dits « domaines de Weiss », ont une réalité physique indiscutable : on les rend visibles au microscope, on les mesure, on fait basculer leur moment magnétique à volonté, on les déforme, etc. Reste à expliquer, bien sûr, le mystérieux champ cristallin qui assure à tous les moments magnétiques d'un même domaine de Weiss un parallélisme rigoureux. Il devait revenir à la mécanique quantique (encore elle!) d'expliquer par des forces d'échange d'origine purement quantique la nature de ce champ cristallin *a priori* mystérieux. On peut dire qu'aujourd'hui le ferromagnétisme est dans l'ensemble assez bien connu et expliqué. Comprenant mieux les substances ferromagnétiques, on a pu construire des aimants permanents, des électro-aimants à performances plus poussées, de prix de revient plus faible, ayant des propriétés plus diversifiées et mieux adaptées aux nombreuses applications. Par exemple, la découverte de solides ferromagnétiques, électriquement isolants, appelés **ferrites,** a engendré des applications remarquables dans le domaine des ondes électromagnétiques d'hyperfréquence.

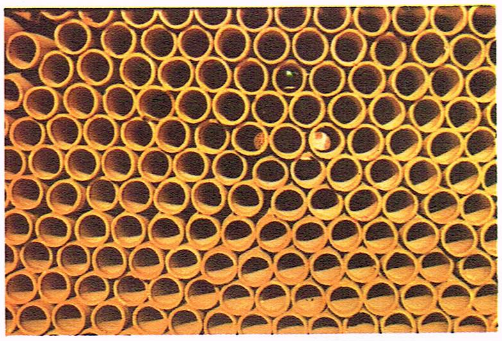
résonance magnétique

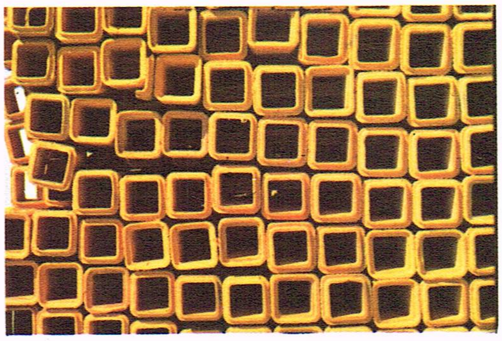
On peut montrer qu'un petit moment magnétique permanent placé dans un champ magnétique $\vec{H}_0$ (de direction et d'intensité fixes) peut tourner autour de



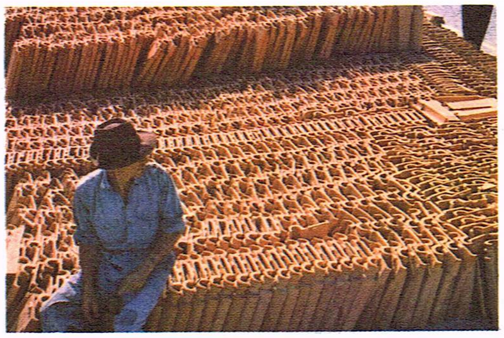
verre et béton

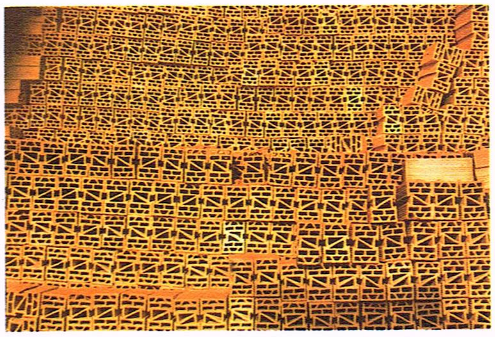


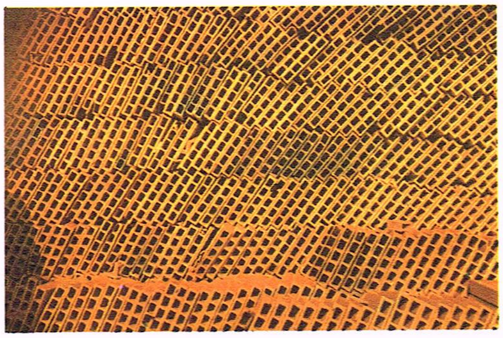


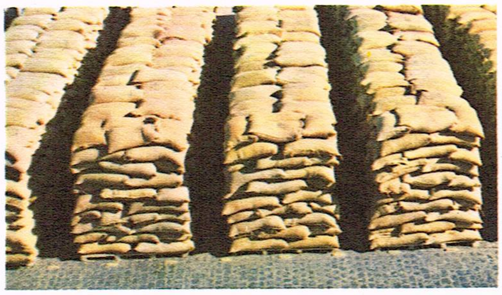


argile

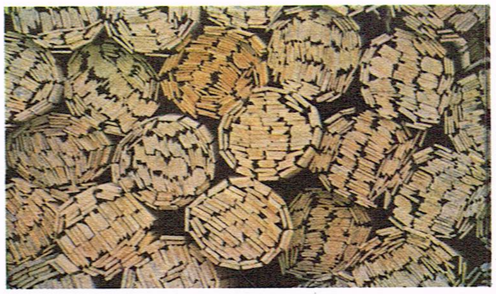








jute



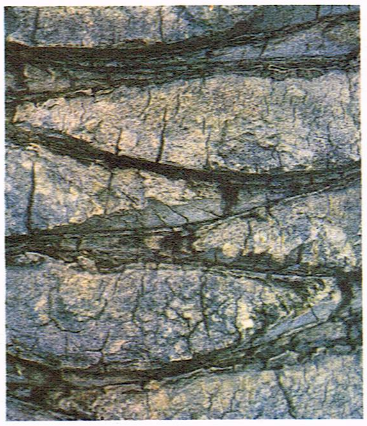
bois



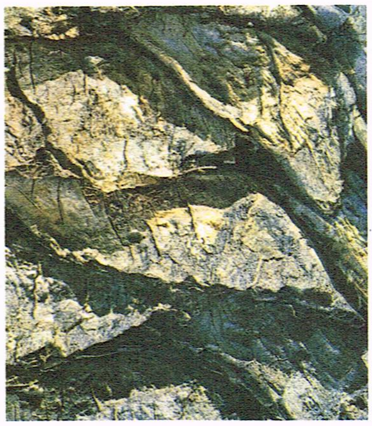
palme



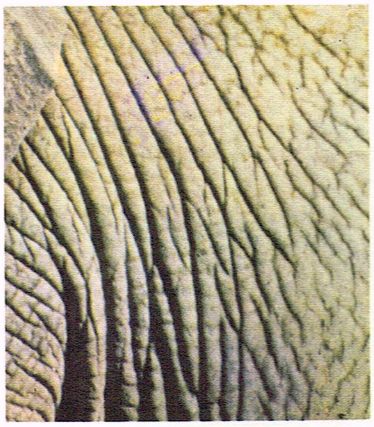
roche



écorce



écorce



derme

Phot. Jidébé.

la direction de ce champ avec une pulsation $\omega_0 = g\mu_0 H_0/\hbar$ ($\hbar$ est la constante de Planck, g un nombre voisin de l'unité, et μ_0 est égal soit à un magnéton de Bohr $[e\hbar/2mc]$, lorsque le moment magnétique est d'origine électronique, soit à un magnéton nucléaire, lorsque le moment magnétique est d'origine nucléaire). En particulier, pour un électron libre, on a $g = 2$, et, par suite, $\omega_0 = \frac{2 H_0}{\hbar} \cdot \frac{e\hbar}{2 mc}$.

On en déduit que la fréquence ($f_0 = \omega_0/2\pi$) vaut en mégacycles 2,80 H_0 (si H_0 est exprimé en œrsteds, c'est-à-dire en unités de champ magnétique). Pour un proton, la fréquence f_0 vaut en kilocycles 4,26 H_0.

En présence d'un champ électromagnétique périodique, de pulsation ω proche de ω_0, le moment magnétique élémentaire tourne autour de la direction du champ $\vec{H}_0$ en empruntant de l'énergie au champ électromagnétique; cette absorption d'énergie est d'autant plus grande que la pulsation ω est plus proche de ω_0; en d'autres termes, on a un phénomène de résonance, qui sera d'autant plus aiguë que le système de moments magnétiques aura moins de pertes, sera moins freiné dans sa rotation. De ce phénomène assez simple, appelé **résonance magnétique (électronique ou nucléaire),** résultent immédiatement quelques conséquences importantes.

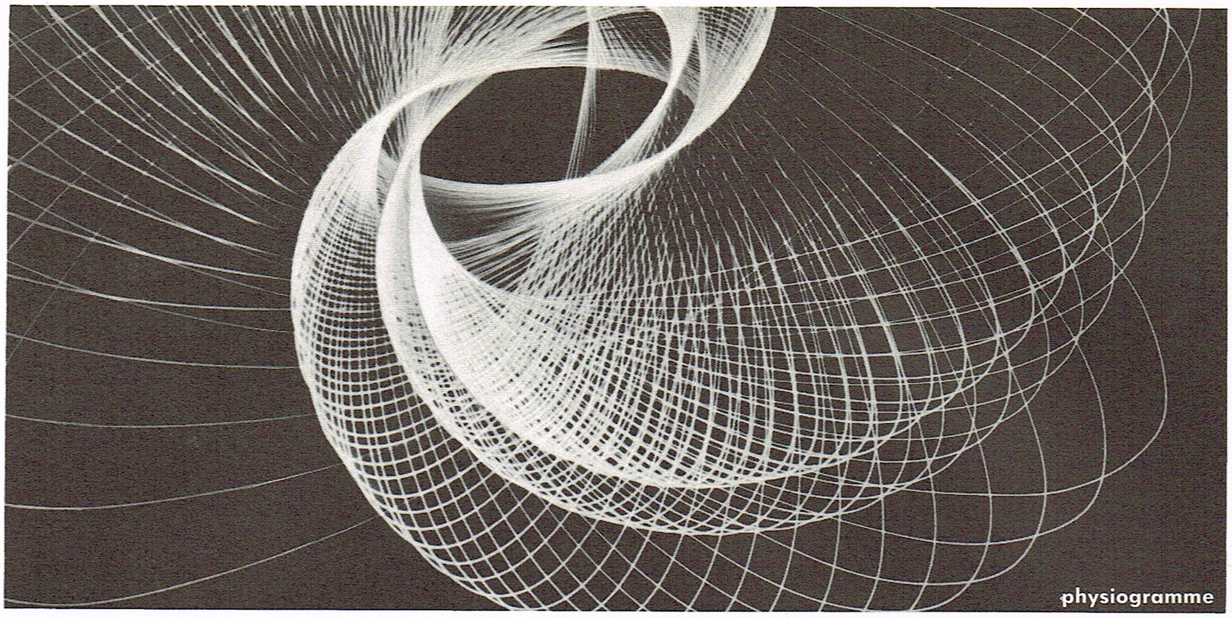
Si l'on détermine avec précision la fréquence de résonance ($f_0 = \omega_0/2\pi$) d'un système de moments magnétiques élémentaires dans un champ magnétique $\vec{H}_0$ connu, on en déduit immédiatement le moment magnétique $g\vec{\mu}_0$ des particules étudiées; d'où une méthode simple, précise, de mesure des moments magnétiques.

Si, au contraire, on fait résonner des particules de moment magnétique connu (des protons, par exemple), la mesure de la fréquence de résonance donne avec précision la valeur du champ magnétique. Enfin, l'étude de la largeur de raie de la résonance en fonction de certains paramètres comme la température, les impuretés du cristal ou la nature du réseau fournit des informations précieuses sur la nature des pertes du système résonnant, c'est-à-dire sur la nature des chocs que subissent les spins et qui viennent troubler la régularité de leur précession autour du champ magnétique fixe $\vec{H}_0$.

Au terme de cette rapide évocation des problèmes de la physique des solides, quelques idées se dégagent.
Tout d'abord, l'étude des solides n'est qu'un chapitre de la physique, un chapitre parmi bien d'autres, consacrés à des phénomènes entre lesquels et les solides il existe les rapports les plus étroits : peut-on parler des propriétés électriques et magnétiques des solides si l'on ignore l'électromagnétisme? Sait-on expliquer la nature hyperfine de la résonance magnétique électronique si l'on néglige les spins nucléaires? Et comment calculer la chaleur spécifique d'un cristal sans la mécanique statistique? Pourtant, la physique des solides a conquis, à défaut d'une indépendance dépourvue de signification, au moins une autonomie interne qui tend à rassembler des faits expérimentaux épars et à unifier des théories auparavant séparées.

Les formalismes de la mécanique quantique construits il y a trente ou trente-cinq ans semblent suffire aux physiciens des solides, même si l'impossibilité pratique de mener à bien certains calculs obscurcit encore à leurs yeux de nombreux problèmes. La physique des solides, physique à notre échelle, ignore les graves problèmes de concepts qui se posent aux deux extrêmes : celui des particules élémentaires de l'infiniment petit et celui de la structure espace-temps-matière de l'infiniment grand.

En somme, pas d'angoisses métaphysiques pour la physique des solides et une échelle humaine à beaucoup d'égards : au contraire de la physique nucléaire et de la physique spatiale, les solides n'exigent pas des milliers de tonnes d'acier et de béton, des centaines de milliers de kilowatts, des équipes de plusieurs centaines de chercheurs; les investissements sont plus modestes et les applications, en général, suivent à quelques années de distance les découvertes du laboratoire. Le rôle de la physique des solides dans l'économie moderne et par suite dans la société future croît sans cesse; l'homme du XXe siècle ne peut plus l'ignorer.



LES MATHÉMATIQUES de nos enfants

par

Jean-Louis RIGAL

**C'est bien plus beau lorsque c'est inutile.
Le beau seul est vrai, le beau seul est utile.**

L'enseignement des mathématiques a subi en quelques années un bouleversement beaucoup plus profond qu'au cours des deux siècles précédents. Et aussi, semble-t-il, la recherche.

Que sont donc ces « mathématiques modernes »? Sont-ce donc encore des mathématiques? Pourquoi un tel vocabulaire, d'apparence si mystérieuse et ésotérique, dans lequel on ne peut se retrouver? Et, si vous êtes physicien, « en quoi tout cela peut-il servir à mieux expliquer le monde de notre expérience? Comment tout cela peut-il s'appliquer »?

Certains lecteurs aussi, pour qui le théorème de Pythagore fut un véritable « pont aux ânes », qui durent souffrir un enseignement dont il était proclamé qu'il fallait avoir la « bosse des maths » pour le comprendre, ne craignent-ils pas d'être définitivement rejetés dans les ténèbres extérieures, dans ce domaine sinistre et de seconde zone, à qui manquera à tout jamais une des plus essentielles clés du monde scientifique? Ou bien, au contraire, un peu de raisonnement et de bon sens, un esprit qui ne se refuse pas aux générations précises — en un mot la raison — suffisent-ils pour la pratiquer couramment et jongler allégrement avec son symbolisme? Tel est bien entendu le cas. Même si l'extrême simplicité des prémisses est compensée par l'extrême technicité des raisonnements. Et s'il est beaucoup plus difficile encore de faire œuvre originale dans la « topologie combinatoire », à partir de la seule connaissance des axiomes et définitions de base, qu'il ne l'est à un élève de sixième, fût-il muni d'un dictionnaire, de traduire Shakespeare dans le texte, du moins pour en comprendre l'esprit et ne pas avoir devant lui l'attitude tremblante du non-initié devant le mage.

Un grand mathématicien nous précise : « Ce qui importe n'est pas tellement de trouver une nouvelle propriété du cercle des neuf points d'un triangle. Ce qui importe beaucoup plus, c'est de connaître la structure même de la géométrie. »

Et voici prononcé le mot clé des mathématiques (prétendument modernes) : structure. *Les mathématiques sont la science des structures (et non celle de l'espace ou celle des nombres, qui, sous les noms de géométrie et d'arithmétique, n'en sont que des branches, et des plus marginales). Et la notion fondamentale qui les domine est celle d'*isomorphisme, *doublée de celle d'*équivalence.

Nous avons déjà donné des exemples où intervenaient efficacement de tels raisonnements et vu à quels rapprochements fructueux ils conduisaient : plusieurs phénomènes physiques (champ newtonien, champ électrique et champ magnétique) satisfont à des équations fondamentales identiques, bien que les variables et les fonctions y figurant aient un contenu physique totalement différent. Ils

Phot. Bouillot.

Nous allons tirer de cette homothétie plusieurs conséquences :

1° Le centre O du cercle circonscrit au triangle ABC est l'orthocentre du triangle *abc*. Il est donc l'homologue de l'orthocentre H du triangle ABC dans l'homothétie $\left(G, -\frac{1}{2}\right)$. *Il en résulte que* O, G, H *sont alignés* et que l'on a :

$$\overrightarrow{GO} = -\frac{1}{2}\overrightarrow{GH}.$$

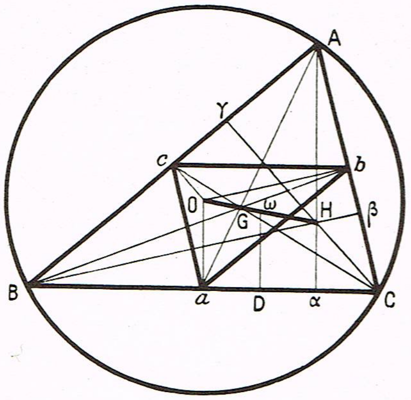


Fig. 154.

Le centre ω du cercle circonscrit au triangle *abc* est l'homologue du point O dans l'homothétie $\left(G, -\frac{1}{2}\right)$.

Le point ω est donc sur la droite OGH et l'on a :

$$\overrightarrow{G\omega} = -\frac{1}{2}\overrightarrow{GO}.$$

Sur la droite OH, *les points se suivent dans l'ordre* OGωH. OG est le tiers de OH, Gω, moitié de OG est le sixième de OH. Il en résulte que ω est au milieu de OH. Ces quatre points forment une division harmonique.

2° Le cercle circonscrit au triangle *abc* est donc le cercle, de centre ω, homologue du cercle circonscrit au triangle ABC dans

52. Propriétés de la multiplication des entiers relatifs.

La multiplication des entiers relatifs possède les propriétés suivantes :

Commutativité. — On a :

$$(a\,;b)\cdot(c\,;d) = (ac + bd\,; bc + ad)$$

et :

$$(c\,;d)\cdot(a\,;b) = (ac + bd\,; bc + ad).$$

On a donc :

$$(a\,;b)\cdot(c\,;d) = (c\,;d)\cdot(a\,;b)$$

ou en posant : $\alpha = (a\,;b)$ et $\beta = (c\,;d)$:

$$\alpha\cdot\beta = \beta\cdot\alpha. \qquad (52\,;1)$$

Associativité. — On vérifie facilement que :

$$[(a\,;b)\cdot(c\,;d)]\cdot(e\,;f) = (a\,;b)\cdot[(c\,;d)\,(e\,;f)]$$

ou :

$$(\alpha.\beta).\gamma = \alpha.(\beta.\gamma). \qquad (52\,;2)$$

Existence d'un élément neutre. -- Cet élément neutre est le nombre relatif $e = (1\,;0) = +1$.

En effet on a:

$$(a\,;b)\cdot(1\,;0) = (1\,;0)\cdot(a;b) = (a\,;b)$$

ou encore :

$$\alpha\cdot e = e\cdot\alpha = \alpha. \qquad (52\,;3)$$

53. Distributivité de la multiplication par rapport à l'addition.

On a :

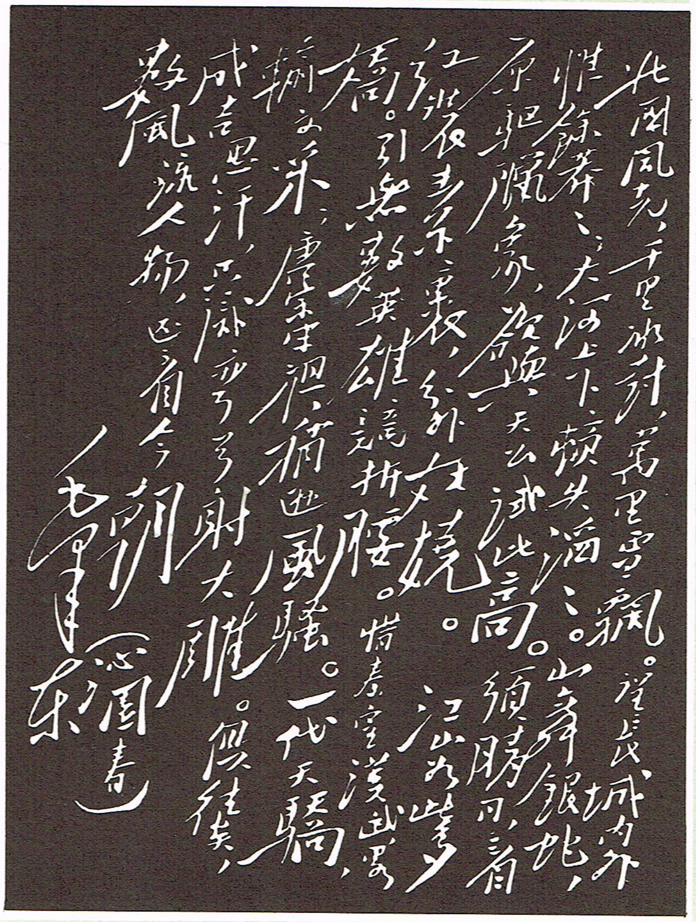
$$(m\,;n)\,[(a\,;b) + (c\,;d)] = (m\,;n)\cdot(a + c\,; b + d)$$
$$= (ma + mc + nb + nd\,; na + nc + mb + md)$$

Un bouleversement dans l'enseignement et dans la recherche.

§ 1. Sous-groupes et groupes quotients de $\mathbf{R}^n$

Introduisons d'abord la convention suivante : si G est un groupe topologique, on a défini, au chap. III (§ 2), le groupe produit G^n de n facteurs égaux à G, pour tout entier $n > 0$; dans ce paragraphe, nous étendrons cette définition au cas où $n = 0$, en convenant que G^0 désignera un groupe réduit à l'élément neutre. Si H est un groupe quelconque, $G^0 \times H$ sera identifié à H.

Sur l'ensemble $\mathbf{R}^n$, nous aurons à considérer, dans ce paragraphe, d'une part sa structure de *groupe topologique* (additif), d'autre part sa structure d'*espace vectoriel* par rapport au corps **R** (chap. VI, § 1, n° 3). Étant donnée une partie A de $\mathbf{R}^n$, nous envisagerons, tantôt le *sous-groupe* de $\mathbf{R}^n$ engendré par A (ensemble des combinaisons linéaires de points de A, à coefficients *entiers*), tantôt le *sous-espace vectoriel* engendré par A (ensemble des combinaisons linéaires de points de A, à coefficients *réels*) ; on aura soin de ne pas confondre ces deux notions. Conformément aux définitions données en Algèbre (*Alg.*, chap. II, 3e éd., § 7, n° 2), nous appellerons *rang* de A la *dimension* du sous-espace vectoriel V engendré par A ; dire que A est de rang p équivaut donc à dire qu'il existe p points $\mathbf{x}_i \in A$, formant un *système libre* par rapport au corps **R** (autrement dit, la relation $\sum_i t_i\mathbf{x}_i = 0$, où les t_i sont *réels*, entraîne $t_i = 0$ pour tout i), et constituant une *base* de V (ce qui veut dire que tout point de V est combinaison linéaire à coefficients réels des $\mathbf{x}_i$).



les mathématiques = du chinois

Il s'agit de deux langues, faciles quand on connaît le vocabulaire. Les mathématiques sont même beaucoup plus universelles et communicables; il suffit d'en savoir le vocabulaire (la sémantique).

donnent lieu à des calculs identiques, se développant parallèlement. Ceux-ci, ne faisant intervenir que les équations et non la nature effective des phénomènes étudiés, aboutissent évidemment à des conclusions formellement identiques, notamment à des définitions parallèles de l'énergie mécanique, de l'énergie électrique et de l'énergie magnétique : une fois faite la théorie d'un des champs, il devient inutile de recommencer tous les calculs pour le second et, plus généralement, pour tout champ satisfaisant aux mêmes lois (ou hypothèses).

Un mathématicien dirait : les deux phénomènes ont, par rapport aux calculs effectués, même structure (donnent lieu à des ensembles isomorphes). Ce qui signifie exactement qu'ils « sont régis par les mêmes lois », plus généralement encore qu'ils « satisfont aux mêmes hypothèses » et sont donc préjudiciables du même formalisme (des mêmes calculs).

Tout revient finalement, ni plus ni moins, à la remarque fondamentale suivante, dont l'évidence et la généralité expliquent l'essentielle importance (non seulement les mathématiques, mais encore toute pensée logique reposent sur elle) :

Des conclusions identiques se déduisent toujours et nécessairement d'hypothèses identiques (il faudrait ajouter la proposition tout aussi essentielle et trop souvent oubliée : les conclusions dépendent des hypothèses).

Cela admis, ne nous cachons pas — encore une fois — que le travail de réflexion logique et d'élaboration structurale des mathématiciens, quoique simple en principe, est, de par cette simplicité même, d'une technicité et d'un niveau d'abstraction exceptionnellement élevés. Nous ne pouvons sérieusement prétendre, surtout en quelques pages, mettre le lecteur au courant des principales structures (on en compte des centaines), ni lui laisser croire qu'il sera rapidement capable de faire œuvre originale dans un domaine où des années d'austères études sont nécessaires.

Qu'y a-t-il d'étonnant à cela? Ne faut-il pas apprendre pendant des années la grammaire et le vocabulaire avant de pouvoir lire Racine ou Rabelais et de jouir pleinement de la richesse et de la beauté de leur langue? La simple connaissance (vite acquise) des règles des échecs suffit certes pour pouvoir lire une partie jouée par un grand maître; mais, plus on acquiert, année par année, d'expérience, mieux on comprend les finesses du jeu, et plus l'admiration pour un Capablanca ou un Alekhine s'accroît.

Il en est de même pour nous : nous espérons faire comprendre comment raisonne un mathématicien, et répondre à ces questions qui nous ont été si souvent posées, toujours avec curiosité, souvent avec une angoisse bien excessive : « Qu'est-ce donc que la recherche mathématique? **Comment peut-on démontrer un nouveau théorème?** » Peut-être, à travers l'effort aride que nous demanderons au lecteur, et malgré lui, passera-t-il un peu de la joie de la découverte qu'éprouve un chercheur! En tout cas, le lecteur se sera familiarisé avec un vocabulaire qui envahit l'ensemble des sciences (les Français sont très en retard à ce sujet) et sera moins choqué lorsqu'il le rencontrera, nécessairement et de plus en plus. En tout cas, la barrière qui sépare le spécialiste du non-mathématicien semblera moins infranchissable. Barrière en partie réelle et normale, dans la mesure où elle provient de la technicité éminente du sujet, mais en grande partie amplifiée par l'habitude que nous avons d'un enseignement qui se prétend mathématique, alors qu'il n'est que géométrique, et même purement euclidien (et ce, bien que la géométrie euclidienne, historiquement la première à avoir été découverte et approfondie il y a plus de vingt siècles par les Grecs, incapables de compter, soit de toutes les structures celle dont la valeur universelle est la moindre, tant en sont complexes et nombreux les axiomes constitutifs).

Nous aurions souhaité montrer au lecteur comment les théories les plus abstraites des mathématiques interviennent tout naturellement et avec beaucoup d'efficacité dans les questions les plus simples et suggèrent des rapprochements aussi profonds qu'inattendus : comment la considération des rotations (géométrie plane) ou celle des restes de la division des polynômes (algèbre) conduisent à la notion de nombres complexes, c'est-à-dire à la vraie théorie des équations du second degré; comment la considération du reste de la division d'un nombre par un diviseur donné (théorie dite « de la congruence ») éclaire définitivement toute la trigonométrie, toute la théorie des polynômes réguliers et toute l'arithmétique des nombres premiers (combien un tel rapprochement apparaît *a priori* inattendu!); comment la considération des partitions d'un ensemble permet finalement les dénombrements simples et corrects sans lesquels on ne peut guère faire de probabilités; comment la considération des groupes de permutation permet de démontrer (E. Galois) qu'une équation algébrique n'est résoluble algébriquement que si son degré est inférieur à 5, théorème qui est loin d'être évident; comment la considération des invariants géométriques (longueurs et angles) débouche sur la relativité, qui devient alors claire et dont le formalisme apparaît comme naturel (invariants du groupe de Lorentz).

Ce qui nous arrête n'est pas la difficulté **(tous les exemples cités, sauf la théorie des groupes de Galois, sont du niveau de la classe de première),** mais le nombre de pages nécessaires à un exposé vivant. Il ne s'agit pas de tout présenter, de tout savoir (les livres et encyclopédies ne manquent plus, fût-ce les manuels scolaires allant de la classe de sixième à la première des lycées), mais d'aller à la découverte d'un mode de pensée, en donnant un ou deux exemples, choisis parmi les plus simples.

de l'isomorphisme sans le savoir

Il n'est pas de notre propos de montrer comment l'humanité a appris à énumérer et à compter — d'abord 1, 2, puis 1, 2, 3, puis beaucoup plus. Le progrès essentiel est intervenu lorsqu'elle a appris (ce que certaines peuplades ne savent pas encore) que l'addition des chameaux, par exemple, s'opère comme celle des tentes, ou encore celle des outres d'eau. De conclusions multipliées telles que :

un chameau + un chameau font deux chameaux,
une tente + une tente font deux tentes,
une outre + une outre font deux outres, etc.,
elle a déduit naturellement, mais au bout de plusieurs millénaires : **1 + 1 = 2.**

Tout cela semble évident, tant nous y sommes habitués. **Et il nous est difficile d'imaginer l'importance et la difficulté d'une démarche intellectuelle que l'humanité n'a du reste pu faire que récemment :** il y a de cela trois à quatre millénaires, sur les cent que compte notre histoire. (Certaines peuplades de l'Amazonie ou du centre de l'Afrique ne sont même pas encore parvenues à ce stade du dénombrement, et emploient encore des mots différents pour indiquer 1 ou 2 selon le type d'objets dont il s'agit.)

L'ensemble des chameaux, l'ensemble des tentes, l'ensemble des outres, etc., sont des ensembles équivalents dont les éléments (chameaux, tentes, outres) sont de nature différente, mais sur lesquels on peut effectuer certaines opérations (additions) ayant identiquement les mêmes propriétés, ce que traduit

la confusion des mots : **deux** chameaux, **deux** tentes, **deux** outres. On dit de tels ensembles qu'ils sont isomorphes (de même forme) par rapport à l'addition.

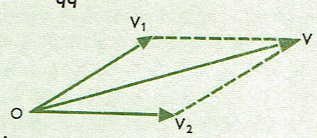
Dès ce stade, auquel tout Français est habitué bien avant d'avoir atteint l'âge de six ans, s'introduisent donc les notions les plus fondamentales des mathématiques : « équivalence » et « isomorphisme ». Laisser croire que de telles notions font partie des « mathématiques modernes » (alors que la géométrie telle qu'on l'apprend encore en seconde, en première ou en mathématiques élémentaires sont des « mathématiques » tout simplement) est un peu abusif, si l'on s'en tient au sens direct du terme. Monsieur Jourdain faisait bien de la prose sans le savoir; le petit écolier fait de l'isomorphisme sans le savoir. Entre sa démarche et celle du plus abstrait des mathématiciens, il y a certes une différence de degré, de technicité, mais non une différence de nature. **L'un et l'autre raisonnent non sur des éléments dont ils connaissent la nature, mais sur des relations qu'ils postulent entre ces éléments** (règles opératoires, règles d'équivalence, règles de calcul), ce qu'on appelle la « structure de l'ensemble considéré ». Est-il besoin de savoir ce que c'est qu'un chameau pour pouvoir affirmer que $1 + 1 = 2$? Bien au contraire, on ne peut l'affirmer que si l'on fait abstraction de la nature physique des chameaux et si l'on ne tient compte que du fait que l'ensemble des chameaux est un ensemble dénombrable, dénombrable voulant strictement dire : sur lequel on peut effectuer des dénombrements.

Différentes définitions, mêmes propriétés.
On vérifie, par exemple, que
$x + y = y + x$ et que $(x + y) + z = x + (y + z)$.

L'addition des entiers : $1 + 1 = 2$.

L'addition des fractions : $\frac{p}{q} + \frac{p'}{q'} = \frac{pq' + p'q}{qq'}$ $(qq' \neq 0)$.

L'addition des vecteurs :
$\vec{V} = \vec{V_1} + \vec{V_2}$.



L'addition des nombres complexes :
$(a, b) + (a', b') = (a + a', b + b')$.

L'addition des polynômes :
$(a_0 x^n + a_1 x^{n-1} + ... + a_n) + (b_0 x^n + ... + b_n) =$
$(a_0 + b_0) x^n + ... + (a_n + b_n)$.

Finalement, on parle de l'addition (et non plus de l'addition des chameaux) et de ses règles (axiomes), dont toutes ses propriétés découlent.

A-t-on observé que l'on peut définir sur les vecteurs une opération ayant les mêmes propriétés constitutives : on parle de l'addition des vecteurs. Il en est de même des additions des fractions, des polynômes, des nombres complexes..., toutes opérations ayant mêmes propriétés, donc même dénomination (« addition »).

Il a suffi de remarquer que les innombrables propriétés de l'addition — qu'il n'est pas trop de longs mois pour apprendre — découlent des axiomes qui la définissent (1); il a suffi de les expliciter pour avoir inventé un concept unifiant, sous le même vocable et sous une même réalité, beaucoup de théories, dont nous n'avons mentionné que certaines, à titre d'exemples.

(1) L'explicitation des axiomes qui sont à la base d'une discipline et le choix du meilleur système d'axiomes sur lequel elle doit être fondée — il y a un grand nombre de possibilités plus ou moins évidemment équivalentes — sont des opérations difficiles. L'axiomatique de la géométrie a été élaborée par cet immense génie des mathématiciens que fut Hilbert (*les Fondements de la géométrie*, 1899).

une relation bien générale : l'égalité

Est-il notion plus générale, plus simple et plus intuitive que celle d'égalité? Cherchons-en donc la signification, et, pour cela, commençons par ouvrir un dictionnaire :

Égalité : qualité des choses égales; relation entre deux choses égales.

Égal : semblable, le même en nature, en quantité, en qualité.

Même : exprime la similitude, l'identité ou la parité.

Identité : ce qui fait qu'une chose est de même nature qu'une autre.

Nous ne serons guère plus avancés en étudiant les diverses définitions où interviennent les mots « égalité » ou « même ». Y a-t-il quelque chose de commun entre eux et n'est-ce pas de l'escroquerie, ou du confusionnisme, que de couvrir plusieurs marchandises sous le même pavillon?

Effectivement, il existe autant de définitions que d'emplois de ce mot passe-partout. **Nul ne peut définir la nature d'une égalité, nul ne peut construire l'égalité : on ne peut construire que « des » égalités.** Mais les mathématiciens se sont aperçus que les relations qu'ils avaient définies entre nombres entiers, entre fractions, entre triangles, entre vecteurs, entre polynômes, entre fonctions avaient plusieurs propriétés communes :

1° Elles s'exercent entre éléments de même nature (deux entiers peuvent être égaux, et de même deux triangles, mais non un entier et un triangle);
2° Tout élément est égal à lui-même (« réflexivité »);
3° Il est équivalent de dire que A est égal à B, ou B égal à A (« symétrie »);
4° Deux éléments égaux à un même troisième sont égaux entre eux (« transitivité »),

toutes propriétés que le lecteur pourra vérifier aisément, avec les définitions que nous avons données en légende des nombres congrus, des fractions égales, des vecteurs équipollents.

Il est tentant de généraliser et d'appeler « égalité » toute relation réflexive, symétrique et transitive, c'est-à-dire telle que :

$$\begin{cases} x = x \text{ (quel que soit } x); \\ (x = y) \text{ est équivalent à } (y = x); \\ (x = y) \text{ et } (y = z) \text{ entraînent } (x = y = z). \end{cases}$$

On est donc passé de constructions d'égalité à la notion de l'égalité, définie par ses propriétés.

Les mathématiques ne sont rien d'autre qu'un tel art : tout le secret, **tout le génie est là, le reste n'est qu'exploitation technique.**

Mais quel intérêt y a-t-il de plaquer un mot sur des notions déjà bien connues? C'est très simple : toutes les conclusions que l'on peut déduire des quatre propriétés de base sont vraies pour toute relation y satisfaisant.

Par exemple : trois (et non plus deux) — et plus généralement même n — éléments égaux à un même élément x sont égaux entre eux :

$$x_1 = x, \quad x_2 = x, \quad x_3 = x, \quad ..., \quad x_n = x;$$
$$(x_5 = x), \ (x = x_3),$$
$$\text{donc} \quad x_3 = x_5 \text{ (transitivité)};$$

et de même pour tout couple (pour toute valeur des indices i et j) :

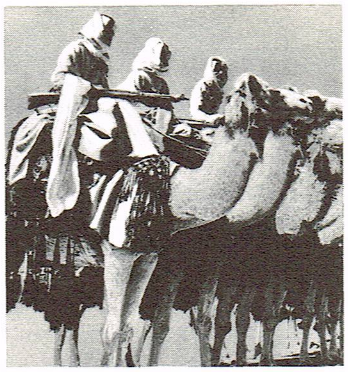
$$x_i = x_j.$$

Deux lignes de démonstration mettent en évidence la raison d'être profonde de conclusions très générales. Et plus n'est besoin de recourir aux opérations coûteuses et absurdes consistant à coller des timbres sur des enveloppes, puis à glisser une lettre dans chaque enveloppe, de coller un timbre sur chaque chameau, etc., pour comparer le nombre de timbres, d'éléments d'ensembles, de chameaux, d'enveloppes et de feuilles de papier

QUELQUES ÉGALITÉS



Le nombre de timbres est égal au nombre d'enveloppes si l'on peut coller un timbre et un seul sur chaque enveloppe, sans qu'il reste à la fin de l'opération ni enveloppe ni timbre inutilisés.



Il y a autant de chameaux que de guerriers si chaque guerrier peut se mettre en selle sur un chameau, sans qu'il reste ni chameau non monté ni guerrier à pied.

nombres égaux

nombres égaux ou congrus (module p)

$x = x' \ (p)$;
si $x - x'$ est un multiple de p,
$x - x' = kp$

fractions égales

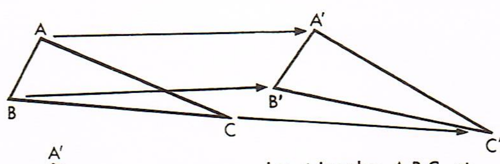
$\frac{1}{2} = \frac{2}{4}$ $(1 \times 4 = 2 \times 2)$;

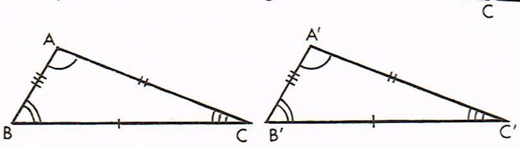
les deux fractions $\frac{p}{q}$ et $\frac{p'}{q'}$ sont égales si, et seulement si $pq' = p'q$.

égalités géométriques

On voit bien qu'elles sont bâties sur la notion de transport, de déplacement. Mais un mystère subsiste : pourquoi deux définitions différentes de l'égalité des triangles? Laquelle est la vraie? Ou bien sont-elles équivalentes?

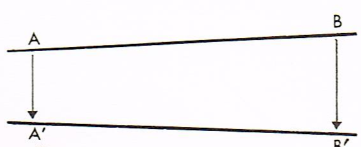
Les triangles A B C et A' B' C' sont égaux si on peut amener A sur A', B sur B' et C sur C'.

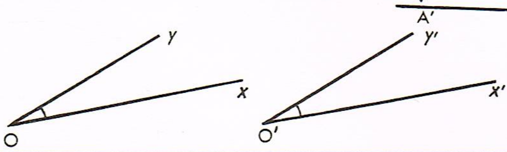




Les triangles A B C et A' B' C' sont égaux si leurs 3 côtés et leurs 3 angles sont égaux deux à deux.

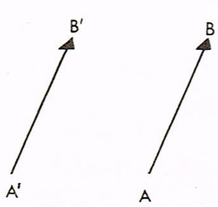
Les longueurs A B et A' B' sont égales si on peut amener A sur A' et B sur B'.





Les angles (Ox, Oy) et (O'x', O'y') sont égaux si on peut amener Ox sur O'x' et Oy sur O'y'.

vecteurs égaux



Deux vecteurs sont égaux (équipollents) s'ils sont de même longueur, parallèles et de même sens.
Deux vecteurs sont égaux si une translation peut amener A sur A' et B sur B'.
Deux vecteurs sont égaux si A B B' A' est un parallélogramme.

polynômes et fonctions égales

$(x + 1)^2 - 2x = x^2 + 1 = (x - 1)^2 + 2x$
$\sin(x + y) = \sin x \cos y + \sin y \cos x$

Deux polynômes sont égaux s'ils ont les mêmes coefficients.
Deux fonctions sont égales si elles prennent les mêmes valeurs pour des valeurs correspondantes de la ou des variables.

que l'on se propose d'étudier. Et il n'est plus besoin d'une autre démonstration pour chaque égalité : une pour l'égalité des entiers, une pour celle des fractions, une pour celle des longueurs...

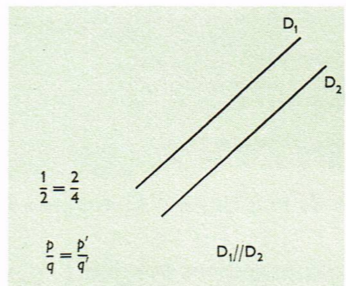
Quelle économie de pensée!

La conclusion était si simple et si évidente qu'il semble que nous ayons employé un marteau-pilon pour écraser une mouche. Cet effet provient de ce que nous n'avons donné que la première étape d'une longue route de raisons — « si simples et si faciles » (Descartes) — aboutissant aux résultats les plus abstraits et les plus difficiles de la théorie des nombres entiers.

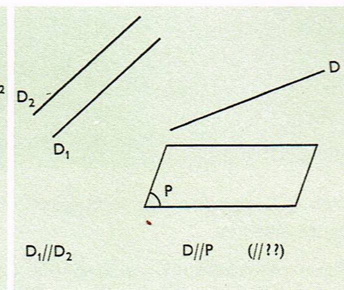
vrai et faux parallélisme

Il est clair que les notions de similitude (pour les triangles), d'équiparité (pour les entiers : congruence module 2), de parallélisme (de droites) sont des relations d'égalité. Ne dit-on pas de deux droites parallèles qu'elles ont même direction?

Remarquons cependant combien pénibles sont les théorèmes usuels concernant le parallélisme. Si l'on définit « parallèle » comme « n'ayant aucun point commun », cette relation n'est pas



Deux notions très voisines



Deux faux frères

une égalité, car elle n'est pas réflexive (une droite n'est pas parallèle à elle-même). Et la raison de se venger : l'énoncé des théorèmes devient très compliqué. Par exemple : deux droites parallèles à une même troisième sont parallèles entre elles ou confondues. On arrive à la notion de « parallèle ou confondue » qui, elle, est bien une relation d'égalité entre droites.

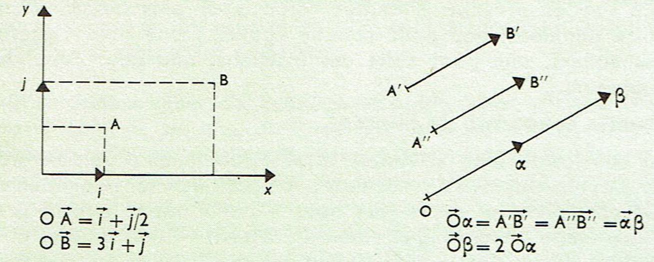
Il y a un abus de langage beaucoup plus grave : certains livres parlent encore de parallélisme de droite et de plan (une droite et un plan sont parallèles s'ils n'ont aucun point commun). Cette relation entre éléments différents (une droite et un plan), n'étant ni réflexive ni transitive, ne peut être une égalité (ce que tout le monde sait, et exprime de manière toute simple : on ne dit jamais qu'une droite et un plan ont la même direction); il y a confusionnisme à couvrir sous le même pavillon deux notions totalement différentes, dont l'une est une égalité, l'autre non, et qui donnent lieu à des théorèmes fondamentalement différents (deux droites parallèles à un même plan ne sont évidemment pas parallèles).

Finalement, la notion d'« aucun point commun » s'est révélée bien décevante et a dû céder la place à celle de « même direction » (c'est-à-dire à une relation d'égalité). **Là est l'essentiel du travail difficile du mathématicien : distinguer les structures fructueuses et générales — rarement évidentes — des structures stériles.** Art difficile, mais nécessité fondamentale.

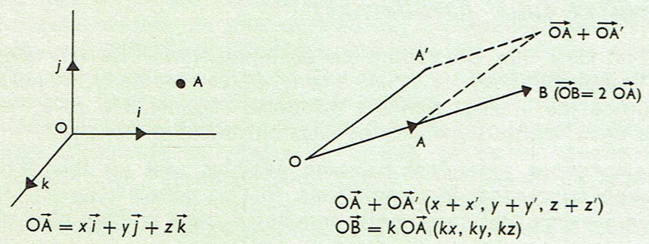
vecteurs et espaces linéaires

On a défini, dans le plan, les vecteurs « liés » comme des couples de deux points dans un ordre bien déterminé : (A, B) ou $\overrightarrow{AB}$; puis les vecteurs libres comme l'ensemble de tous les vecteurs liés égaux entre eux; un vecteur libre se représente donc par

Phot. Larousse, Keystone.



Espace vectoriel à deux dimensions



Espace vectoriel à trois dimensions

$\vec{OA}\,(x, y, z, t, u, v, w...)$ $\quad$ $OA + OA'\,(x + x', y + y', z + z', t + t', u + u', v + v', w + w')$
$k\vec{OA}\ (kx, ky, kz, kt, ku, kv, kw)$

Espace vectoriel à sept dimensions

l'un quelconque des vecteurs liés de la « classe d'équivalence » ainsi définie.

Pour ces vecteurs, on a défini deux opérations possibles :

a) la somme $\vec{AB} + \vec{BC} = \vec{AC}$ (on peut toujours amener le second vecteur à avoir pour origine l'extrémité du premier);

b) le produit par un nombre : $k \cdot \vec{AB}$;

et le résultat est indépendant du choix des « représentants ».

A partir de ces deux seules opérations, on peut poursuivre très loin l'étude, notamment voir qu'il existe deux vecteurs $\vec{i}$ et $\vec{j}$ portés par Ox et par Oy, tels que tout vecteur puisse se mettre d'une manière et d'une seule sous la forme $\vec{OA} = x\vec{i} + y\vec{i}$ (et de même $\vec{OB} = x'\vec{i} + y'\vec{j}$).

Les nombres x et y sont dits les « composantes du vecteur $\vec{OA}$ » (et de même x', y' pour $\vec{OB}$). La somme se définit par la somme des composantes, et le produit $k\vec{OA}$ par le produit des composantes par k :

$\vec{OA} + \vec{OB}$ a pour composante $(x + x', y + y')$;
$k\vec{OA}$ a pour composante (kx, ky).

Même théorie pour l'espace à trois dimensions, sauf qu'il faut maintenant trois coordonnées (x, y, z) pour définir un vecteur OA (x, y, z).

Mais, alors, pourquoi ne pas considérer des ensembles de 4, 7 nombres coordonnés, ou même d'une infinité? Toute la théorie subsiste dans ses conséquences immédiates ou lointaines. Même si un espace à quatre ou sept dimensions n'est plus représentable géométriquement (en fait, même un espace à trois dimensions n'est pas représentable sur le papier : il faut le projeter), il a des propriétés identiques à celles du plan : il n'est en rien plus mystérieux, comme une mauvaise littérature a voulu le faire croire. **Dire qu'un espace a *n* dimensions ne signifie rien de plus que : il faut se donner *n* paramètres pour fixer la position d'un point.** Par exemple, un événement est bien déterminé par sa localisation (abscisse, ordonnée et cote, par exemple) et sa date, soit quatre paramètres.

les opérateurs et leurs produits

Mots bien mystérieux, et pourtant réalités bien simples.

Elever un nombre au carré est une opération qui fait correspondre à un nombre x un autre (non négatif) x^2 selon un processus bien défini (on multiplie le nombre par lui-même). De même, faire tourner de 100 grades $\left(\frac{\pi}{2}\text{ radians}\right)$ un point autour d'un axe fait correspondre à un point M un autre point M' en vertu d'un processus bien défini. On peut également faire correspondre à un vecteur un nombre : sa longueur.

On pourrait multiplier les exemples : quand à un élément d'un ensemble (nombre, point) on fait correspondre un élément d'un autre ensemble (qui peut être confondu avec le premier), on dit que l'on applique un opérateur et l'on écrit :

$x \to x^2$ (opérateur : élévation au carré);
$M \to M' = R(M)$ [opérateur : rotation de 100 gr autour de Δ];
$AB \to$ distance AB (opérateur : mesure de la longueur);
$x \to f(x)$ [notation générale d'un opérateur].

Il est très fréquent que l'on ait à faire de nouvelles opérations sur le résultat $f(x)$. Par exemple, $2x^2$ est obtenu en élevant x au carré, puis en multipliant le résultat par 2 :

$$x \xrightarrow[\text{(élévation au carré)}]{} x^2 \xrightarrow[\text{(multiplication par 2)}]{} 2x^2.$$

Plus généralement, le résultat final

$$x \underset{(f)}{\to} [f(x)] \underset{(g)}{\to} g[f(x)]$$

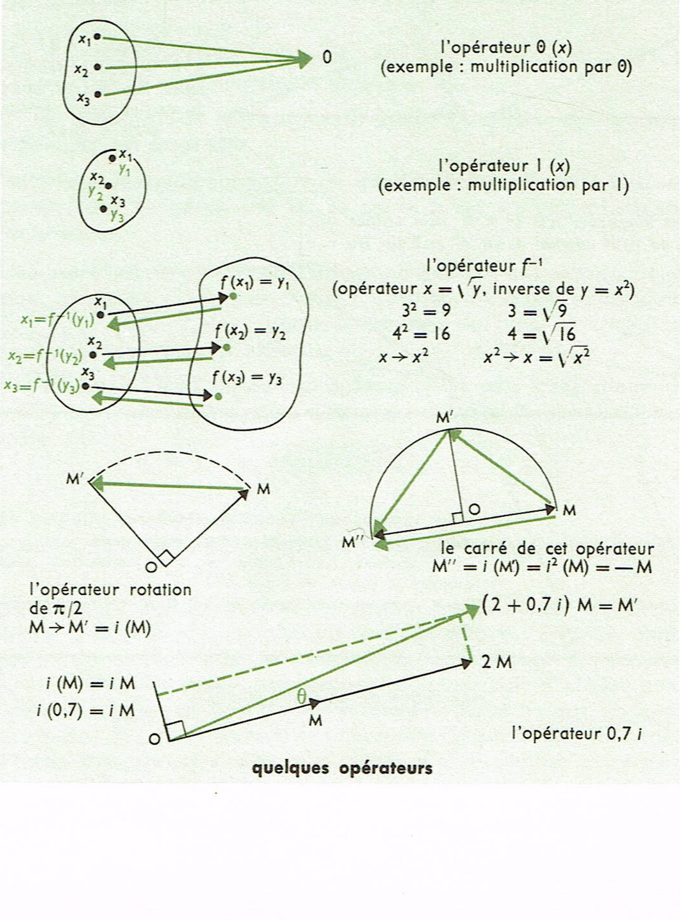
sera tout naturellement désigné par $g[f(x)]$:

$$x \to [gf(x)],$$

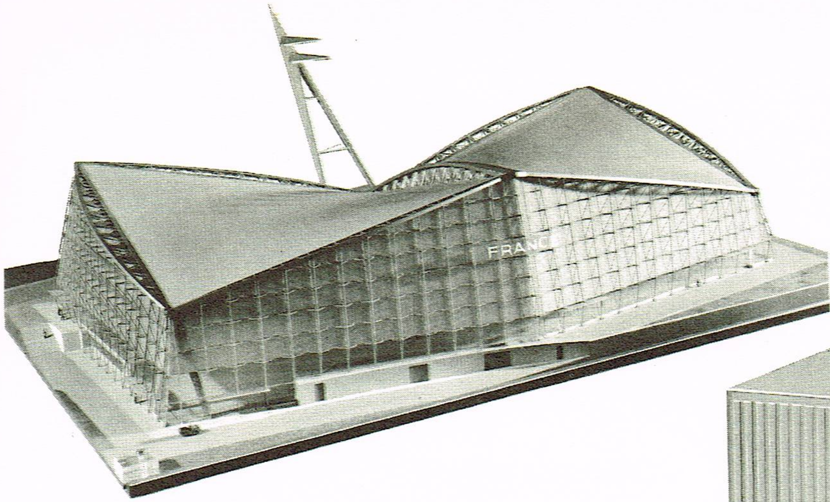
que l'on préfère écrire plus simplement $gf(x)$,
et l'on dit que l'on a effectué le produit (1) gf des opérateurs f et g.

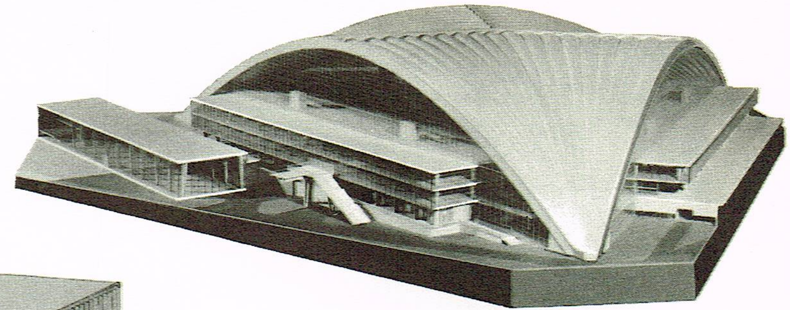
---

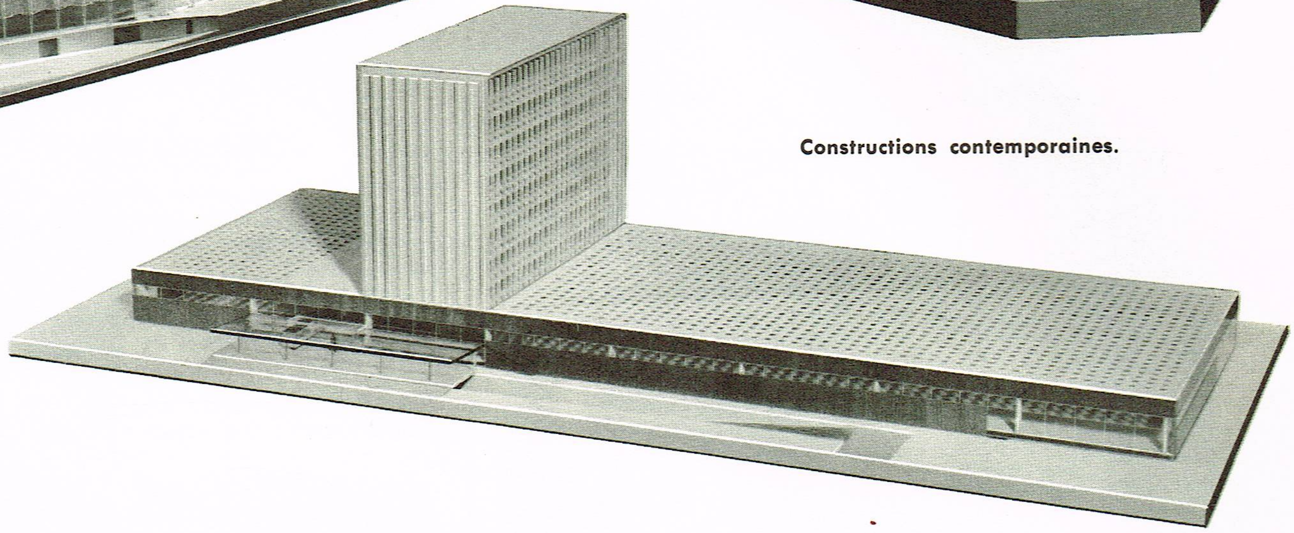
(1) Remarquons que l'on est conduit naturellement à écrire le produit gf, alors que l'on effectue d'abord f, puis g; comme à classer les périodiques de gauche à droite dans sa bibliothèque, si l'on veut que les pages se suivent dans l'ordre naturel.



quelques opérateurs







Constructions contemporaines.

dans ces maquettes d'architectes, beauté et qualité grâce aux mathématiques

Il est en effet remarquable que l'on retrouve toutes les propriétés constitutives du produit. Soit Θ, l'opérateur (nul) qui à tout élément x fait correspondre l'élément nul de l'ensemble image :

$$x \to \Theta(x) = 0;$$

soit I l'opérateur (identité) :

$x \to I(x) = x$ (ensemble objet et ensemble image sont alors nécessairement identiques);

et f^{-1} l'opérateur (inverse) :

$$f(x) \to f^{-1}[f(x)] = x,$$

c'est-à-dire l'opérateur faisant passer, non de x à $f(x)$, mais de $f(x)$ à x. On retrouve facilement les relations classiques du produit :

$\Theta \cdot f = \Theta$	$(x \to f(x) \to 0)$;	$(0x = 0)$;
$I \cdot f = f$	$(x \to f[x] \to f[x])$;	$(1x = x)$;
$f^{-1}f = I$	$(x \to f(x) \to x)$;	$(1/x.x = 1)$;
$h(gf) = (hg)f$	$(x \to h\{g[f(x)]\})$;	$c(ba) = (cb)a$.

A une exception près notable, le produit n'est pas commutatif, c'est-à-dire

$gf \neq fg$ $\qquad ab = ba$

Il n'est pas indifférent de faire d'abord f, puis g, ou le contraire :

$$(2x)^2 \neq 2(x^2).$$

Les produits d'opérateurs satisfont donc à certaines règles de l'algèbre, non à toutes; on dit qu'ils forment un groupe non commutatif.

Ayant défini le produit de deux opérateurs, on peut définir le carré, ou toute puissance, d'un opérateur tel que la rotation d'angle droit.

Faire n rotations d'angles droits revient à faire une seule rotation d'angles « n droits ».

En appelant maintenant i cet opérateur « rotation d'un angle droit autour de O », on trouve la formule (que nous généraliserons plus tard)

$$i^m i^n = i^{m+n}$$

(l'angle final est la somme des angles des deux rotations), et notamment :

$$i^2 = -1$$

(une rotation d'angle deux droits est une symétrie),

$$i^4 = 1$$

(une rotation d'angle quatre droits est une identité). Remarquons que cette dernière formule est en accord avec les deux précédentes :

$$i^4 = i^2(i^2) = -1^2 = 1.$$

de l'algèbre des similitudes à celle des nombres complexes...

On peut définir d'autres règles sur les opérateurs, comme va nous le montrer l'exemple suivant.

Considérons deux transformateurs (on appelle « transformateur » tout opérateur portant sur des points ou sur des figures géométriques) :

$$M \to m = f(M),$$
$$M \to m' = f'(M),$$

et soit m'' la somme des points telle que

$$m'' = m + m' = f(M) + f'(M) = f''(M).$$

On dira naturellement que f'' est la somme de f et de f'.

Nous allons prendre un exemple plus concret, et qui parlera plus aux géomètres, en considérant les deux transformateurs simples :

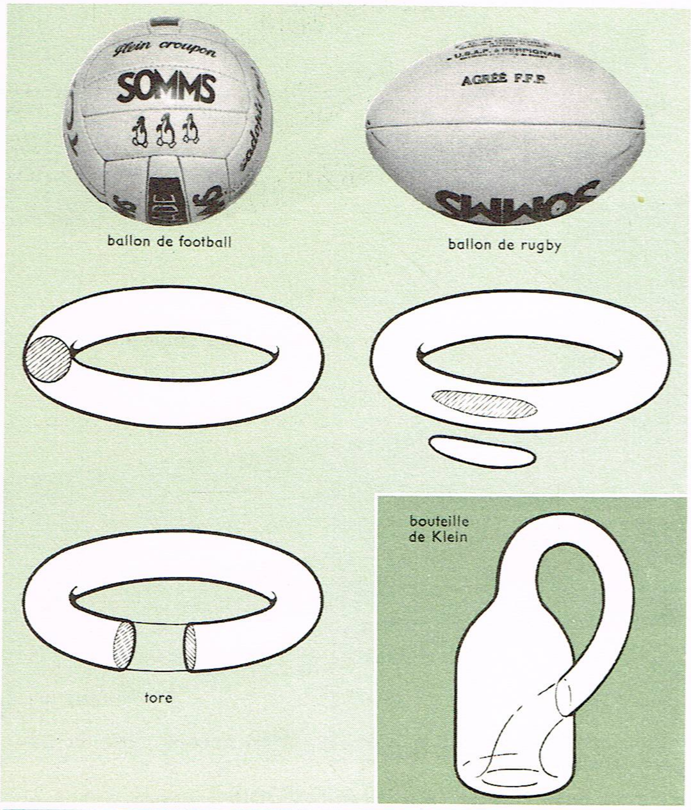
$M \to k(M) = kM$ (homothétie de rapport k);
$M \to i(M) = iM$ (rotation déjà considérée de 100 grades autour de O).

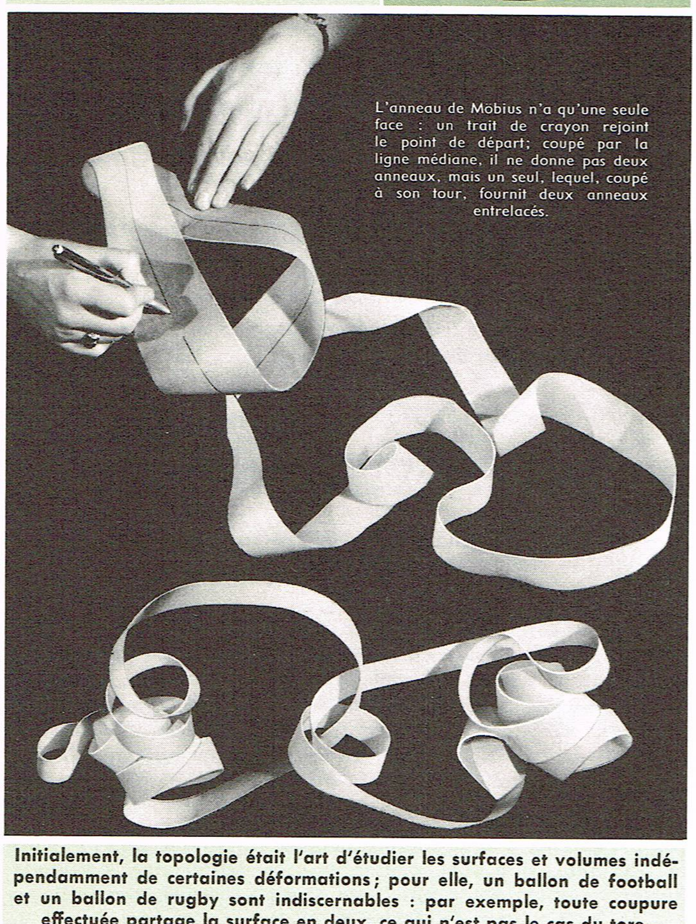
L'opérateur $0{,}7i$ sera par définition obtenu en faisant tourner M d'un angle droit, puis en faisant l'homothétie de rapport 0,7 sur le point obtenu. Quant à l'opérateur somme $2 + 0{,}7i$, il conserve les angles (invariants de la transformation). Or, conserver les angles est le propre des similitudes (déplacements suivis d'une homothétie; ou, en termes plus mathématiques : produits d'une rotation d'angle θ par une homothétie de rapport k).

D'où il résulte, d'une manière peut-être indirecte, mais efficace et profonde, que tout couple $a + bi$ définit une similitude. Réciproquement, toute similitude peut s'exprimer sous la forme $a + bi$, a et b étant l'abscisse et l'ordonnée du transformé du point A_0 ($x = 1$, $y = 0$).

Finalement, le produit de deux similitudes est une similitude (conservation des angles), et l'on démontre (nous nous excusons

Phot. Larousse.





L'anneau de Möbius n'a qu'une seule face : un trait de crayon rejoint le point de départ; coupé par la ligne médiane, il ne donne pas deux anneaux, mais un seul, lequel, coupé à son tour, fournit deux anneaux entrelacés.

Initialement, la topologie était l'art d'étudier les surfaces et volumes indépendamment de certaines déformations; pour elle, un ballon de football et un ballon de rugby sont indiscernables : par exemple, toute coupure effectuée partage la surface en deux, ce qui n'est pas le cas du tore. L'anneau de Möbius (et la bouteille de Klein) sont des curiosités amusantes. Récemment, le domaine de la topologie, devenue étude des voisinages, a vu son champ d'activité envahir presque toute la mathématique moderne, dont elle constitue un fleuron essentiel avec la théorie des ensembles.

de ne pas le faire ici : chaque étape est simple, mais leur ensemble long et fastidieux; et il ne peut être question ni de lasser le lecteur avec une telle technique ni de développer en deux pages ce que l'élève de l'enseignement secondaire met quelques mois à comprendre) que l'algèbre des similitudes est identique à celle des nombres réels, avec la possibilité supplémentaire de remplacer i^2 par — 1. Par exemple :

$(a + bc)(c + di) = ac + i(ad + bc) + i^2bd = (ac - bd) + i(ad + bc)$;
$(a + bi)(a - bi) = a^2 - i^2b^2 = a^2 + b^2$,

et par conséquent on calcule ainsi l'inverse d'un nombre complexe :

$$\frac{1}{a + bi} = \frac{a - b}{a^2 + b^2} = \frac{a}{a^2 + b^2} - i\frac{b}{a^2 + b^2}.$$

De même, nous avons déjà vu comment s'introduisait une fonction exponentielle (opérateur : rotation d'angle φ) généralisant l'opérateur i (rotation d'angle droit)

$$\cos\varphi + i\sin\varphi = e^{i\varphi} \qquad \left[e^{i\frac{\pi}{2}} = 1\right],$$

jouissant de propriétés comparables à celles déjà définies pour les réels, et notamment de

$$e^{i\varphi}e^{i\varphi'} = e^{i(\varphi + \varphi')},$$

formule qui signifie tout simplement que le produit de deux rotations d'angles φ et φ' autour de O est une rotation d'angle $\varphi + \varphi'$.

Deux énoncés d'apparence différente, mais en fait voisins :

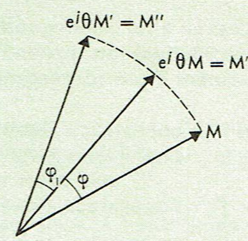
Le produit de deux rotations est la rotation d'angle comme

$$R(\varphi)\,R(\varphi_1) = R(\varphi + \varphi_1).$$

Le produit de deux exponentielles est l'exponentielle de la somme :

$$e^{x} \cdot e^{x'} = e^{x + x'}.$$

le symbolisme $R(\varphi) = e^{i}\varphi$ explicite cette profonde parenté.



$$e^{i\pi} = -1$$

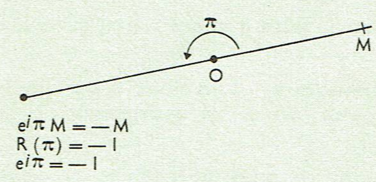
La formule de Moivre et ses applications; admirons qu'une formule simple relie des nombres aussi divers que :

$$e = 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \dots + \frac{1}{n!} + \dots = 2{,}718\,281\,828\,46\dots$$

(rapport de la circonférence au diamètre d'un cercle = 3,141 592 653 59...)

$$i = \sqrt{-1} \ (!).$$

Elle signifie simplement qu'une rotation de deux droits est une symétrie !



Tout cela n'est-il pas évident et ne s'agit-il que de désigner de manière étrange et artificielle ce que tout le monde sait, de manière à le rendre inintelligible?

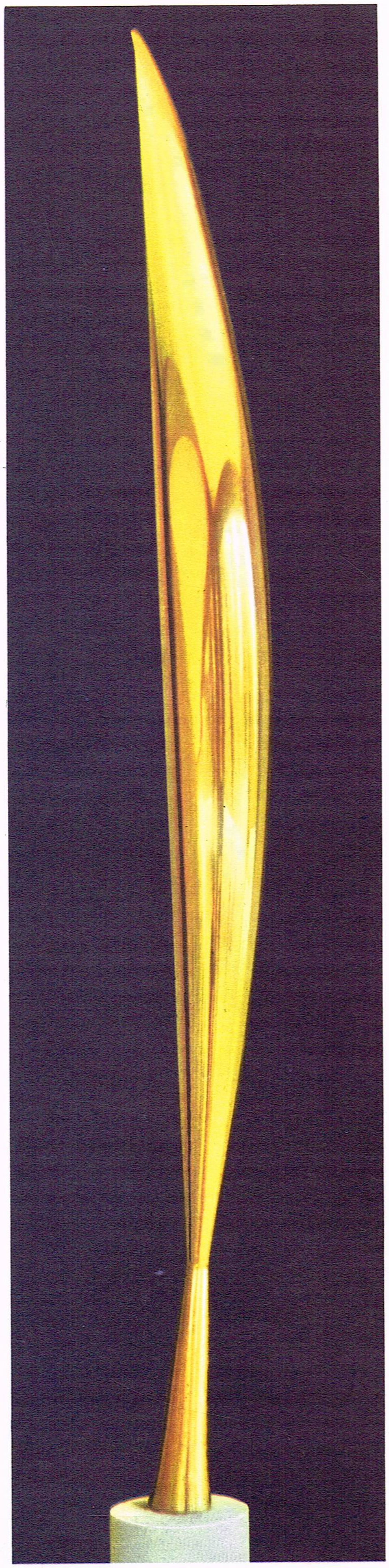
Non certes. Il suffit de développer la dernière formule pour que notre longue démarche rapporte enfin ses premiers dividendes :

$$\begin{aligned} e^{i(\varphi + \varphi')} &= \cos(\varphi + \varphi') + i\sin(\varphi + \varphi') \\ &= (\cos\varphi + i\sin\varphi)(\cos\varphi' + i\sin\varphi') \\ &= \cos\varphi\cos\varphi' - \sin\varphi\sin\varphi' + i(\cos\varphi\sin\varphi' + \sin\varphi\cos\varphi'), \end{aligned}$$

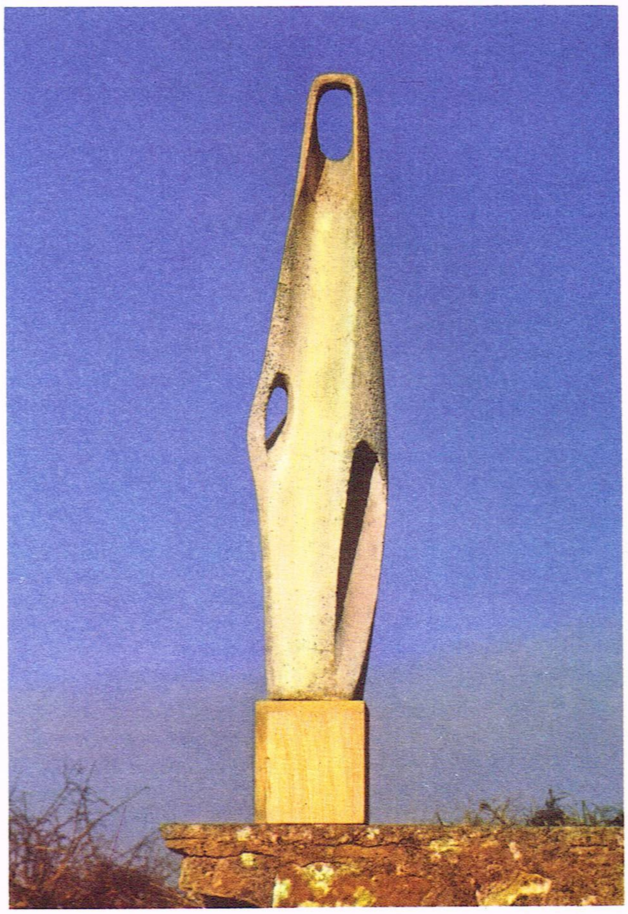
ce qui donne tout de suite la valeur des lignes trigonométriques de la somme $\varphi + \varphi'$:

$$\begin{cases} \cos(\varphi + \varphi') = \cos\varphi\cos\varphi' - \sin\varphi\sin\varphi', \\ \sin(\varphi + \varphi') = \cos\varphi\sin\varphi' + \sin\varphi\cos\varphi'. \end{cases}$$

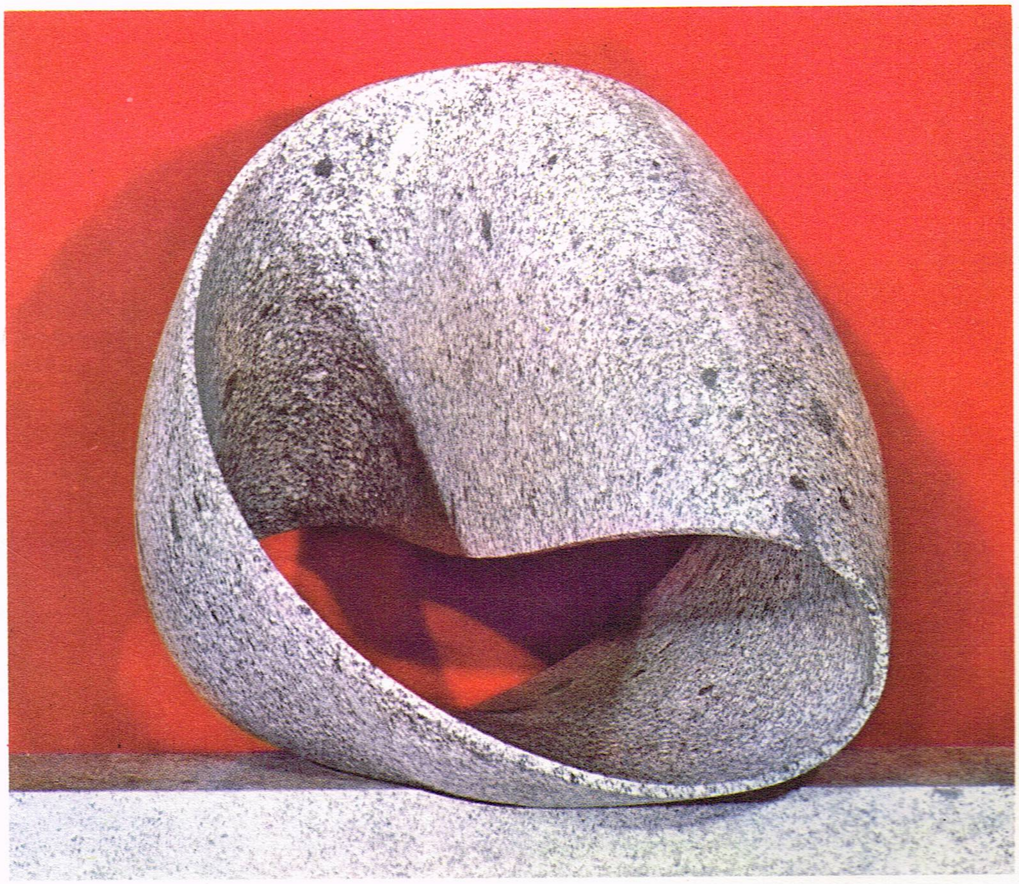
La sobriété
des mathématiques
et la pureté
des figures
qu'elles étudient
n'ont pas été
sans influencer
l'art et la pensée
modernes.



« L'Oiseau d'or »
par Constantin Brâncusi



« Vierge », par Gilles
Coll. part. (Doc. Rigal).



« Le Ruban sans fin » (mathématiquement parlant : un *anneau de Möebius*), par Max Bill, *musée national d'Art moderne. (Phot. Larousse.)*

De même, la formule évidente (définition d'une puissance)

$$(e^{i\varphi})^n = (e^{in\varphi})$$

s'écrit (formule de Moivre)

$$(\cos\varphi + i\sin\varphi)^n = \cos n\varphi + i\sin n\varphi$$

et fournit la valeur des lignes trigonométriques de l'angle $n\varphi$.

Par exemple,
$(\cos 3\varphi + i\sin 3\varphi) = (\cos\varphi + i\sin\varphi)^3$
$= \cos^3\varphi + 3i\cos^2\varphi\sin\varphi - 3\cos\varphi\sin^2\varphi - i\sin^3\varphi$
donne
$\cos 3\varphi = \cos^3\varphi - 3\cos\varphi\sin^2\varphi;$
$\sin 3\varphi = 3\cos^2\varphi\sin\varphi - \sin^3\varphi.$

Combien la trigonométrie a pu nous ennuyer, à juste titre, dans notre jeunesse! Il a pourtant suffi de faire un petit détour par la théorie des similitudes, sans craindre d'aborder l'étude des nombres complexes, pour redécouvrir une formule (inventée en 1730, il y a donc plus de deux siècles) qui rend clairs beaucoup de calculs, fastidieux si l'on n'en saisit pas la raison profonde.

Nous avons donc construit géométriquement des êtres, dépendant de deux paramètres (c'est-à-dire définis par un couple de nombres réels, ou par la position d'un point dans le plan), généralisant les nombres habituels et obéissant aux mêmes lois (addition, soustraction, multiplication et division), et dont un au moins a pour carré l'opérateur — 1 isomorphe au nombre — 1.

Cela est toujours une construction; il nous faut donc aller plus loin, vers la démarche essentielle du mathématicien : construire les axiomes et les structures qui correspondent aux êtres qui n'ont servi que comme support à notre intuition. On aboutit aux axiomes ci-dessous, définissant les « nombres complexes ».

Il s'agit de couples (1) de deux nombres réels (a, b), jouissant des propriétés suivantes (que le lecteur pourra vérifier en les écrivant $[a + bi]$ et en leur appliquant les lois usuelles de l'algèbre) :

1° $(a, b) = (c, d)$ signifie $a = c, b = d$;
2° $(a, b) + (a', b') = (a + a', b + b')$;
3° $(a, b) - (a', b') = (a - a', b - b')$;
4° $(a, b)(c, d) = (ac - bd, ad + bc)$;
5° $(a, b)^{-1} = (a/a^2 + b^2 - b/a^2 + b^2)$;
6° (a, o) représente le nombre réel a.

On vérifie notamment que
$(0,1)^2 = (0,1).(0,1) = (-1,0)$ représentant — 1 (formules 4 et 6).

(1) Il n'est pas plus étonnant de définir un complexe comme un couple de réels qu'une fraction rationnelle comme un couple d'entiers. Même un nombre entier relatif (positif, négatif ou nul) est défini au mieux comme un couple d'entiers positifs (n_1, n_2), isomorphe à $n_1 - n_2$ si $n_1 > n_2$, à $-(n_2 - n_1)$ si $n_1 < n_2$, $(n_{1,0})$ étant isomorphe à l'entier n_1.

...et de celle des couples complexes à celle des équations algébriques

Que d'ennui et que d'ennuis aussi, dans la résolution d'une équation du second degré, qui introduit une curieuse quantité $b^2 - 4ac$, dont l'examen inaugure généralement une pénible discussion, comportant trois cas possibles! Et tout cela parce qu'un nombre négatif ne pouvait être considéré comme le carré d'un nombre réel.

Une fois cet obstacle levé de la façon suivante

$$\sqrt{-k^2} = ik \text{ (puisque } [ik]^2 = -k^2 \leqslant 0),$$

plus rien ne s'oppose à écrire qu'une équation du second degré **($a \neq 0$)** a toujours deux racines (exceptionnellement égales) satisfaisant toujours (qu'elles soient « réelles » ou complexes) aux mêmes relations :

$$x_1 = -\frac{b + \sqrt{b^2 - 4ac}}{2a}, \qquad x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a},$$

$$x_1 + x_2 = \frac{-b}{a}, \qquad x_1 x_2 = \frac{c}{a}.$$

Plus généralement, un théorème dû au « prince des mathématiciens », l'Allemand Gauss (1777-1855), affirme que toute équation algébrique de degré n

$$P(x) = x^n + a_1 x^{n-1} + \ldots + a_n = 0$$

a exactement n solutions (distinctes ou non), ce qui permet d'écrire

$$P(x) = (x - x_1)(x - x_2)\ldots(x - x_n),$$

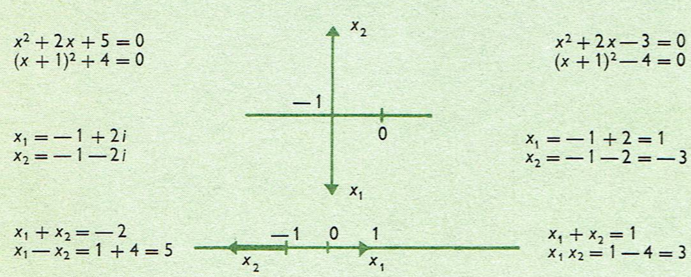
et, en identifiant les deux expressions ci-dessus, de généraliser les formules précédentes et les rendre valables quel que soit le degré, quelle que soit la nature des coefficients et des solutions :
$\boldsymbol{\Sigma x_i = -a_1}$ (somme des racines);
$\boldsymbol{x_1 x_2 \ldots x_n = (-1)^n a_n}$ (produit des racines).

La théorie des fonctions de variables complexes — qu'il ne peut être question d'aborder ici — est encore plus riche, puisqu'une même fonction peut être considérée comme fonction d'une seule variable complexe $(x + iy)$ ou du couple de deux variables réelles (x, y).

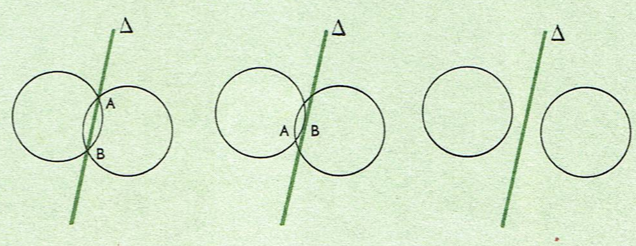
vers des ensembles de nombres de plus en plus riches

Les mathématiciens ont d'abord connu les entiers positifs $(0, 1, 2, \ldots, n, \ldots)$.

C'est un ensemble intéressant, mais bien limité : on peut faire la somme de deux éléments, ou le produit, mais non la sous-



deux équations analogues et non pas différentes



trois cas de figures donnant lieu à la même théorie de l'axe radical, même si l'on ne peut représenter A et B dans le troisième cas

L'introduction d'éléments complexes permet de considérer sous la même rubrique la théorie de l'intersection des courbes algébriques : une droite et un cercle se coupent en 2 points; deux cercles, en 2 points (à distance finie), qui définissent leur axe radical, droite toujours réelle, même si les points d'intersection des cercles ne le sont pas (principe de Poncelet).

traction (3 — 4 n'est pas un entier positif), ni la division (il n'y a aucun entier qui, multiplié par 5, donne 3).

Pour remédier à cette pauvreté, on a imaginé de généraliser la notion de nombre, c'est-à-dire de trouver un ensemble plus riche jouissant de propriétés plus étendues (mais comprenant l'ensemble des entiers).

Ce fut alors l'invention des entiers négatifs, entre lesquels on peut définir trois opérations : l'addition, la soustraction et la multiplication, avec les règles bien connues (on dit que les entiers forment un anneau). Puis ce fut l'invention des nombres rationnels, c'est-à-dire des couples d'entiers (p, q), soumis à la seule restriction $q \neq 0$. Les « quatre » opérations sont possibles : les rationnels forment un corps, qui comprend comme éléments particuliers les entiers (on les écrit plus commodément p/q).

On eut l'idée de considérer d'autres opérations, comme l'extraction de la racine carrée, et l'on vit rapidement que ce n'était pas possible. En effet :

a) $\sqrt{2}$ ne peut être rationnel (aucun nombre rationnel ne peut avoir 2 comme carré) [1]; de même π ne peut être rationnel (ni même solution d'aucune équation algébrique à coefficients entiers). Une difficile généralisation aboutit à la notion de nombres réels, généralisant encore celle de nombres rationnels (qui en sont des cas particuliers), dont ils possèdent les propriétés (leur définition fait appel à la difficile notion de suites infinies, ou à celle de tri effectué sur tous les rationnels. Ils en possèdent encore d'autres, nouvelles : tout point d'une droite est représentable par un nombre réel, l'ensemble en est continu; tout nombre positif a une racine carrée;...

b) $\sqrt{-1}$ ne peut exister. On a vu par quel détour on a résolu le problème en considérant des couples (a, b). De tels couples furent d'abord appelés « imaginaires » — tant on avait du mal à concevoir des nombres de carrés négatifs —, puis « complexes »; appellation beaucoup plus satisfaisante, car elle mettait en valeur la richesse de ces nombres définis par un couple de nombres réels (a, b). **De même qu'un rationnel est un couple d'entiers (p, q), un complexe est un couple de deux nombres réels et l'ensemble des nombres de la forme (a, o) est purement assimilé (isomorphe) à celui des nombres réels a.**

Du même coup, cette dernière généralisation (prolongement) résolvait de nombreux problèmes, ainsi que le montrait le théorème de Gauss. Il ne semblait guère nécessaire d'aller plus loin.

C'était ne rien comprendre à l'inlassable curiosité des mathématiciens que de s'imaginer qu'ils ne chercheraient pas à généraliser encore. Mais ils purent démontrer qu'il est impossible d'inventer des nombres encore plus généraux jouissant de toutes les propriétés des nombres réels (déjà la notion d'ordre disparaît entre les nombres complexes, et aucune relation du type $[a, b] \leqslant [a', b']$ ne peut exister, tout en possédant les propriétés de la relation d'ordre entre réels). De telles généralisations de la notion de nombre existent (quaternions, ou couples de quatre nombres a, b, c, d; matrices ou tableaux rectangulaires), mais sans qu'aucun possède — ni puisse posséder — toutes les propriétés des réels. Par exemple, ni le produit des quaternions ni celui des matrices n'est commutatif; il existe des matrices non nulles n'ayant pas d'inverse...

Ainsi sont atteintes les limites de la notion de nombres, et nous voyons comment « la notion de nombre pur ne fut acquise par l'homme qu'au prix d'immenses difficultés : le caractère abstrait présenté de nos jours par les mathématiques résulte d'une très longue suite de généralisations » (F. Foutan).

(1) Théorème facile à montrer de manière amusante : s'il existe une fraction irréductible (p, q) de carré égal à 2, on peut écrire $p^2 = 2q^2$, d'où il résulte que p^2 et donc que p est pair. En écrivant $p = 2p'$, il vient $2p'^2 = q^2$, ce qui démontre que q est pair aussi. Cela est en contradiction avec l'hypothèse de l'irréductibilité de la fraction (p, q), qui interdit que p et q aient 2 comme diviseur commun. La pensée grecque buta irrémédiablement sur ce scandale, affreux pour elle.

nombres entiers positifs

0, 1, 2...

$a + b$, $a \cdot b$ sont entiers

$a - b$, a/b ne le sont pas nécessairement

nombres entiers relatifs

..., ..., — 2, — 1, 0, 1, 2...

$a + b$, $a - b$, $a \times b$ le sont encore

mais pas nécessairement a/b [exactement, les entiers relatifs sont définis comme couples d'entiers (a, b), satisfaisant aux règles
$(a, b) \rightarrow a - b$ $(a \geqslant b)$
$(a, 0) \rightarrow a$ $(b = 0)$
$(a, b) + (c, d) = (a + c, b + d)$...
(on complétera facilement)]

nombres rationnels

$(p, q) = [q \neq 0]$

$(p, q)\,(p', q') = (pp', qq')$
$(p, q) + (p', q') = (pq' + p'q, qq')$
$(p, q)^{-1} = (q, p)$
les 4 opérations sont possibles. $(p, 1)$ correspond à un entier

nombres réels

$\pi = 3{,}141\,59...$
$\sqrt{2} = 1{,}414...$
(une infinité de décimales entières)

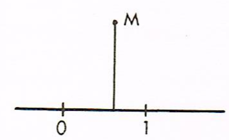
exactement, $\sqrt{2}$ est la limite de la suite infinie des nombres
1
1,4
1,41
1,414...

O ——— M
$x = OM$

les nombres réels permettent de représenter les points d'une droite

nombres complexes

$0{,}7 + 2i$



les nombres complexes permettent de représenter les points du plan et de résoudre les polynômes. Ils satisfont aux règles
$(a, 0) \rightarrow a$
$(a, b) + (c, d) = (a + c, b + d)$
$(a, b)\,(c, d) = (ac - bd, ad + bc)$
$(a, b)^{-1} = \left(\dfrac{a}{a^2 + b^2}, \dfrac{-b}{a^2 + b^2}\right)$

quaternions

$(a + bj + ck + dl)$

$a + bj + 0k + 0l$ est isomorphe au nombre complexe
$(a, b, 0, 0) \rightarrow (a, b)$
$(a, 0, 0, 0) \rightarrow (a)$

les quaternions permettent de représenter les espaces de 4 dimensions, mais leur produit n'est pas commutatif
$(a, b, c, d)\,(a', b', c', d') \neq (a', b', c', d')\,(a, b, c, d)$

Par exemple :
$jk = -kj = l$
$kl = -lk = j$
$lj = -jl = k$

les matrices

(tableaux de nombres réels) représentent des transformations linéaires, mais leur produit n'est pas non plus commutatif

$$\begin{Vmatrix} 1 & 0 & 2 \\ 4 & 3 & 5 \\ 6 & 8 & 0 \end{Vmatrix} \times \begin{Vmatrix} 0 & 1 & 1 \\ 3 & 5 & 2 \\ 1 & 4 & 7 \end{Vmatrix} = \begin{Vmatrix} 2 & 9 & 15 \\ 14 & 39 & 45 \\ 24 & 46 & 22 \end{Vmatrix}$$

$$\begin{Vmatrix} 0 & 1 & 1 \\ 3 & 5 & 5 \\ 1 & 4 & 7 \end{Vmatrix} \times \begin{Vmatrix} 1 & 0 & 2 \\ 4 & 3 & 5 \\ 6 & 8 & 0 \end{Vmatrix} = \begin{Vmatrix} 10 & 11 & 5 \\ 53 & 55 & 31 \\ 59 & 68 & 22 \end{Vmatrix}$$

les matrices (1, 1) $\| k \|$ sont isomorphes à un réel (k)

les matrices particulières (2, 2) $\begin{Vmatrix} a & b \\ -b & a \end{Vmatrix}$ correspondent aux complexes (a, b)

Généralisation successive de la notion des nombres

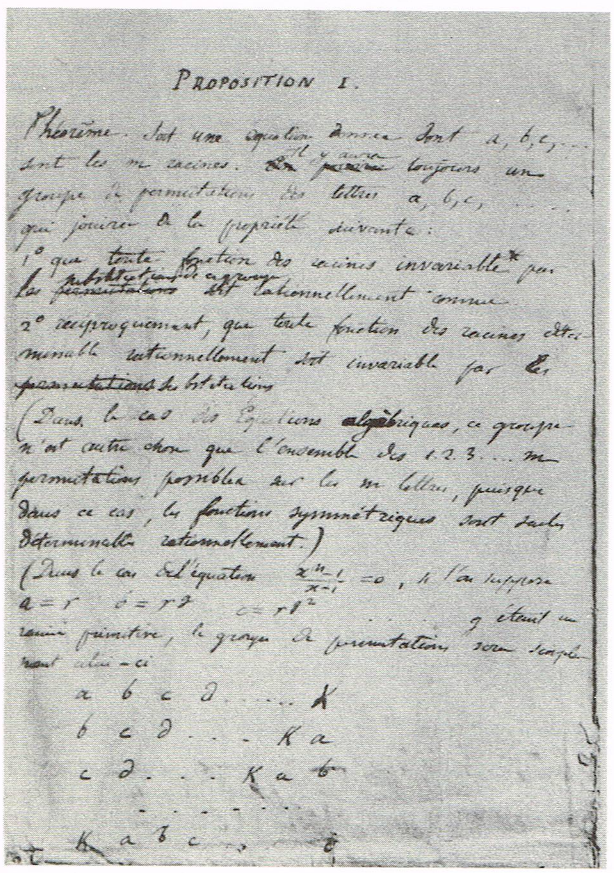
les mathématiques sont-elles utiles?

Fort bien, direz-vous, je comprends que l'on se délecte dans de telles activités. Mais cela sert-il? Pourquoi? Et comment? Comment redescendre sur terre après de telles rêveries?

Savoir pourquoi le monde est intelligible, et pourquoi la réflexion logique a prise sur lui et confère à l'homme puissance et efficacité, dépasserait largement notre propos. Aussi nous contenterons-nous de donner quelques exemples révélateurs où l'on verra la pensée la plus abstraite et la plus désintéressée déboucher plus ou moins rapidement sur l'action.

Deux siècles et demi avant la naissance du Christ, un philosophe grec, disciple d'Archimède, Apollonios de Perga, étudie les propriétés des sections coniques (section d'un cône de révolution par un plan) et découvre, entre autres, le théorème suivant : « L'aire du parallélogramme construit sur deux demi-diamètres conjugués est constante et égale au produit des demi-axes de l'ellipse. » Les Grecs ne s'intéressaient qu'à la beauté formelle des théorèmes et leurs spéculations philosophiques se suffisaient à elles-mêmes. L'idée, caractérisant notre siècle, que la pensée abstraite pouvait et devait déboucher sur l'action leur était totalement étrangère (sauf, peut-être, à Archimède).

C'est cependant cette découverte qui permit à Newton, vingt siècles plus tard, de s'apercevoir que les célèbres lois de Kepler expliquaient le mouvement des planètes en termes d'une accélération centrale en r^{-2} (loi de l'attraction universelle). **Un beau théorème (Apollonios), une patiente observation (loi de Kepler), un génie pour établir un extraordinaire rapprochement entre la théorie et l'observation (Newton), voici les facteurs typiques de la genèse d'une loi scientifique.**


PROPOSITION I.

Théorème. Soit une équation donnée dont a, b, c, ... sont les m racines. Il y aura toujours un groupe de permutations des lettres a, b, c, ... qui jouira de la propriété suivante :

1° que toute fonction des racines invariable par les substitutions de ce groupe soit rationnellement connue

2° réciproquement, que toute fonction des racines déterminable rationnellement soit invariable par les substitutions

(Dans le cas des équations algébriques, ce groupe n'est autre chose que l'ensemble des 1.2.3. ... m permutations possibles sur les m lettres, puisque dans ce cas, les fonctions symétriques sont seules déterminables rationnellement.)

(Dans le cas de l'équation $\frac{x^n-1}{x-1}=0$, si l'on suppose $a=r$ $b=r^g$ $c=r^{g^2}$ g étant une racine primitive, le groupe de permutation sera simplement celui-ci

a b c d ..... k

b c d .. k a

c d . . k a b

k a b c ...

Une page du manuscrit d'Evariste Galois. Archives de l'Académie des sciences.

Le jeune Evariste Galois (1811-1832) crée, avant de mourir assassiné dans un duel à vingt et un ans, la théorie des groupes (il est tellement en avance sur son siècle qu'il a fallu une génération pour commencer à percevoir la puissance insurpassable de cette notion, et plusieurs pour l'expliciter pleinement). Non seulement cette théorie permet de démontrer que l'équation algébrique la plus générale de degré supérieur à quatre ne peut avoir de solution calculable algébriquement (ce qui n'a pas encore l'air d'un résultat essentiel), mais encore elle permet de comprendre la raison d'être profonde de l'organisation des cristaux, et c'est sur elle qu'est fondée toute la cristallographie.

Le logicien britannique Boole (1815-1864) étudie les lois les plus abstraites de la théorie des propositions logiques. Un siècle plus tard, on s'aperçoit que l'algèbre de Boole est à la base même de la théorie des calculateurs électroniques : une des plus belles inventions de ces dernières années s'appuie sur une théorie si élaborée qu'elle semblait gratuite (et n'avait intéressé personne à l'époque), ainsi que sur l'étude physique des lampes et des semi-conducteurs.

Riemann (1826-1866) fait une théorie des géométries non euclidiennes, fondées sur des axiomes tels que : **« Par deux points passent deux droites (ou, exceptionnellement, une infinité). » « La somme des angles d'un triangle est égale à deux droits, plus l'aire de ce triangle. »** Sans une construction « aussi visiblement aberrante » la relativité n'aurait pas vu le jour quelque trente ans plus tard.

Le théoricien allemand David Hilbert (1862-1943), déjà mentionné pour l'axiomatique qu'il a développée pour la géométrie euclidienne, fait la théorie des espaces très généraux et abstraits (espaces linéaires complexes à une infinité de dimensions), qui se révèle très rapidement, sous l'impulsion énergique et géniale de son inventeur, être l'outil le plus efficace de toute la physique théorique, outil dont les applications ne se comptent plus. **C'est exactement un an après la publication de ses théories que de Broglie fonde précisément la physique moderne.**

Ce palmarès, que l'on pourrait continuer presque indéfiniment, est révélateur de la puissance de ce langage concis et précis que sont les mathématiques.

Il n'est jamais facile de savoir si telle théorie donnera lieu à des applications profondes, à plus ou moins longue échéance, ou si, au contraire, elle tombera définitivement dans l'oubli. Et les mathématiciens se défendent de juger selon un tel critère leurs travaux. Les physiciens de 1980 auront-ils précisément besoin de la théorie dite de la cohomologie, ou de celle des algèbriques simples de Malcev? Nul ne peut encore le dire.

En revanche, on peut et on doit affirmer que la science est une aventure totale nécessitant les constructions de tous les hommes — des plus abstraites aux plus expérimentales. Et que le monde de la civilisation de nos fils sortira autant des élaborations les plus gratuites des mathématiciens que de l'observation patiente et ordonnée des physiciens, ou des nécessités de la technologie.

Et ne serait-ce que parce que les mathématiques fournissent le langage le plus précis qui soit nécessaire à la communication universelle et sans ambiguïté d'un certain type de réflexion.

Phot. Larousse.

COMMUNAUTÉ des SCIENTIFIQUES

Pour aborder les quelques réflexions générales qui fermeront cet ouvrage, tournons nos regards vers les quatre pages bordées de vert de la mécanique quantique. Elles nous ont procuré bien des sujets de réflexion : qui intéressent-elles, par qui peuvent-elles être lues et comprises, ces quatre pages sur les mystérieux "opérateurs" et sur le "spin" ? Ne pouvait-on pas les supprimer tout simplement ? Non, ce n'eût pas été honnête, car les opérateurs sont en fait le "b-a ba" de la mécanique quantique; c'est la première clé qui ouvre la porte de tout le formalisme, qui permet de s'introduire dans cet immense et merveilleux palais de la description des atomes. On l'apprend d'ailleurs couramment à notre époque : la notion d'"opérateur" est devenue classique dans l'enseignement des sciences, et non seulement dans le "supérieur", dans la préparation des grandes écoles, mais également dans les centres intellectuellement plus modestes ; les jeunes techniciens qui suivent les cours des Arts et Métiers ou des cours similaires et qui désirent posséder un bagage convenable pour leur vie professionnelle se familiarisent avec ce langage abstrait, et ils se comptent par milliers.

Cela nous conduit à mesurer la vitesse de renouvellement des connaissances à notre époque, car il ne s'agit - dans ces quatre pages bordées de vert - que de notions assez élémentaires pour la jeunesse actuelle ; les développements constructifs, ceux qui aboutissent à la prévision des molécules, des constituants chimiques, des choix à opérer pour les semi-conducteurs, à la classification et aux propriétés des particules, sont infiniment plus difficiles et compliqués.

Ainsi, les anciens perdent pied rapidement. Jusque vers trente-cinq ans l'on est bien dans la course, allégrement, en phase avec les nouveaux langages ; pendant les dix années suivantes, on commence à peiner pour suivre, après quoi, progressivement, on prend un retard croissant et l'on ne peut plus rester en phase avec le progrès ; c'est alors que l'on devient un personnage plus administratif, organisant ou encore enseignant les bases classiques : en bref, on est rejeté du mouvement de la recherche. Souvent, le désir de retrouver sa jeunesse intellectuelle saisit l'ancien : il assiste, en élève, sur les bancs des écoliers, aux exposés des jeunes, sans en bien comprendre toutes les implications et toute la profondeur. J'ai vu, au CERN par exemple, pendant un congrès sur les particules, les anciens les plus célèbres, prix Nobel de physique parfois, ceux qui - il y a dix ans - tenaient la vedette, ceux dont chaque conférence était religieusement suivie, écouter humblement en prenant des notes les brillants exposés des nouveaux jeunes champions.

Nous savons que notre connaissance est très limitée, que la nature pose sans cesse des problèmes plus complexes, inattendus, aux multiples faces et que, lorsqu'ils auront été tant bien que mal résolus, d'autres problèmes, insoupçonnés maintenant, se poseront alors. Il n'y a pas à regretter cette évolution rapide : elle nous met bien à notre place. Elle exige de l'imagination, de la patience, un travail constructif, sérieux, solide, continu ; elle requiert impérieusement la constitution de groupes harmonieux et puissants, humainement et amicalement unis pour s'atteler aux problèmes complexes qui se posent ; elle nous pousse à nous développer merveilleusement et nous apprend à la fois la fierté de notre potentiel et une profonde humilité.

Car toute cette science demande infiniment de qualités et de sérieux : elle ne peut se faire que dans l'honnêteté la plus rigoureuse. On se rend mal compte, quand on

n'y a pas participé, à quel point une expérience sur les particules fondamentales, par exemple, exige d'honnêteté scrupuleuse. Publier trop vite, annoncer une découverte incomplètement assurée, sans avoir contrôlé tous les biais éventuels, sans avoir chiffré rigoureusement toutes les erreurs possibles ni envisagé les probabilités les plus minimes des autres explications, et le groupe correspondant est vite contré par d'autres, sa valeur mise en doute : il court à un échec dont il ne se relèvera pas. Il en est de même pour la réalisation d'un photomultiplicateur poussé, d'un transistor, d'un accélérateur : que de travail et d'honnêteté pour obtenir une réalisation valable qui fasse progresser la science ou son application.

Ainsi va le monde scientifique. Il y a lieu de s'en réjouir. Une telle accumulation de travail sérieux, continu, honnête, associant patience et imagination créatrice, effectué selon les mêmes lois, les mêmes règles par tous les hommes attelés à ce travail quelles que soient leurs options philosophiques, c'est certainement un bien qui s'accroît de façon continue dans le capital de notre humanité. L'on se demande souvent si l'humanité est meilleure actuellement, si elle progresse, ou au contraire si elle ne régresse pas à mesure que le progrès technique et scientifique se développe. L'homme est toujours l'homme. Il peut toujours choisir entre le bien ou le mal ; le mal qui est à sa portée est plus grave parfois, plus lourd de conséquences qu'il n'était autrefois. C'est vrai, mais également le bien. Et la discussion peut s'orienter dans un sens ou dans l'autre. Mais pendant que vit le monde, pendant qu'il évolue, pendant que chacun des hommes exerce son choix particulier, un immense travail se développe, dont le volume croît constamment, et vient nourrir, alimenter l'humanité dans son ensemble, apportant un poids croissant pour le bien dans la balance du bien et du mal. En ce sens, il y a progrès par cette accumulation de vertus, par ce développement des meilleures virtualités, sans lesquelles la science ne peut avancer.

Ce qui peut-être frappe le plus celui qui observe la communauté des scientifiques - car elle est bien réelle cette communauté -, c'est une foi dans le progrès, une assurance que, malgré les fluctuations de situations et les épreuves de la vie de chacun, de la vie des peuples, le monde progresse, que son potentiel positif s'accroît, grâce en particulier à l'immense effort commun des scientifiques et des techniciens, qui s'inscrit dans la direction la plus authentique de l'évolution de l'humanité. Il ne s'agit pas d'une confiance aveugle et un peu simpliste dans les bienfaits des applications de la science, mais d'une certitude plutôt que d'une espérance, la certitude d'une participation à une grande œuvre, belle au plein sens du terme.

Ceux qui ont mission de philosopher sur le monde devraient tenir le plus grand compte de cette aspiration scientifique de l'humanité, levier puissant de la jeunesse universelle : il nous semble que l'insuffisance de certaines philosophies existentialistes, qui tiennent compte seulement de l'incidence des applications scientifiques sur le comportement humain, est en grande partie due à la méconnaissance de cette orientation fondamentale de l'être moderne.

Et voici que, évoquant de nouveau les applications, nous sommes amenés à préciser, pour clore ces réflexions, leur place au sein de la pensée scientifique. Nous avons tenu, dans cet ouvrage, à en présenter quelques-unes parmi les plus récentes et les plus spectaculaires. Nous avons cherché à montrer que l'on retrouve des méthodes communes, des volumes industriels du même ordre de grandeur, des complexités similaires pour le développement des deux volets du diptyque. Les grands ordinateurs électroniques, avec leurs équipes de programmeurs, se retrouvent partout. Ils sont indispensables à la connaissance de l'infiniment petit et de l'infiniment grand, au

CERN et à Nançay, tout comme ils le sont à l'énergie nucléaire, aux chemins de fer, voire aux compagnies d'assurances.

Mais malgré le tronc commun des technologies utilisées, la pensée, elle, est très différenciée. Un physicien n'est pas un chimiste, ni un biologiste, et même un physicien des particules est très loin, dans son intimité cérébrale, dans son comportement, d'un physicien de l'état solide ou des très basses températures. Cela est si vrai que la communication est très difficile entre eux et le devient de plus en plus. Par ses structures et ses implications profondes, la science est de moins en moins communicable.

Revenons à l'attitude du scientifique devant les applications : l'homme de science qui étudie les particules ou la radio-astronomie ne les recherche pas. Il faut d'ailleurs bien définir le sens de cette proposition : tout scientifique est attentif à l'utilisation qui sera faite de la science qu'il développe, et cette attention fait partie du comportement humain du savant. La prise de possession du monde le préoccupe et de graves problèmes de conscience se posent devant les utilisations dont la puissance bénéfique ou maléfique peut être considérable. On l'a bien vu avec les physiciens du noyau, Einstein et Oppenheimer en tête, au cours des vingt-cinq dernières années. Mais la recherche fondamentale est bien "fondamentale" ; au CERN, nous étudions pour découvrir la complexité de la nature, la finesse des nouvelles structures subnucléoniques que personne n'a encore pu observer ni découvrir. C'est cette magnifique activité qui requiert toutes les conditions et les qualités dont il vient d'être question.

Lorsque l'on a découvert les isotopes, c'était de la pure connaissance, celle d'une complexité nouvelle des atomes, et nul ne pensait alors à la bombe atomique et aux usines de Pierrelatte. Lorsque nous étudions les antiprotons et leurs annihilations, c'est la connaissance pure qui nous préoccupe. Lorsque l'on étudie les astres éloignés, les émissions radio des galaxies, c'est elle aussi qui est en jeu. Plus tard, les utilisations pratiques apparaîtront, plus tard la compréhension des plasmas stellaires permettra de concevoir de nouvelles possibilités d'énergies sur la Terre. Ce sont d'autres hommes, physiciens et ingénieurs, qui chercheront les applications.

Et même, que dire de leur aspect bénéfique ou maléfique ? Cela dépend de tant de variables et en premier lieu de l'époque envisagée : dix ans ou trente ans après la découverte des neutrons, les utilisations de ces neutrons auront des caractères fort différents ; certaines verront leur intérêt diminué, d'autres se développeront très rapidement, donnant naissance à des technologies insoupçonnables. Mais nous ne pouvons entrer dans le détail de cette analyse complexe : qu'il nous suffise de jeter un coup d'œil sur les développements industriels, médicaux, biologiques des radio-éléments, pour nous convaincre de l'impossibilité, au moment d'une découverte, de prévoir les applications et leurs évolutions dans le temps.

Et puis, les problèmes économiques et surtout politiques dominent les plus puissantes de ces utilisations. L'épée de Damoclès de la bombe atomique est suspendue au-dessus de l'humanité depuis vingt ans et les bombes se perfectionnent et se multiplient chaque jour ! C'est à la sagesse des nations, des plus grandes en particulier, c'est aux ententes entre pays, voire entre continents, que reviennent finalement les grandes décisions : ce n'est plus alors du savant qu'il s'agit, c'est de l'homme en définitive.

Louis LEPRINCE-RINGUET

index

Les chiffres en italique renvoient aux illustrations et les chiffres en gras au paragraphe principal.

Q

R

S

T

U

V

W, Y, Z

Imprimerie LAROUSSE, 1 à 9, rue d'Arcueil, Montrouge (Seine).
Février 1965. — Dépôt légal 1965-1er. — N° 3 175. — N° série éditeur 3 223. — IMPRIMÉ EN FRANCE *(Printed in France)*. — 16 120 A-9-65.

nouvelle collection

LAROUSSE

de large vulgarisation scientifique

POUR CONNAÎTRE

Une équipe internationale de savants dirige cette nouvelle collection Larousse qui constitue la plus attrayante et la plus facile des initiations à toutes les connaissances indispensables de notre époque.

Chaque volume est conçu selon la même méthode originale :

— le texte est associé à d'innombrables illustrations en noir et en couleurs (croquis, schémas, photographies, cartes, dessins de Hans Erni...) ;

— chaque double page traite d'un sujet particulier ;

— un glossaire des termes scientifiques et un index permettent d'utiliser ces encyclopédies comme des dictionnaires.

LES SCIENCES

chimie - physique - astronomie

LA NATURE

la Terre - les plantes - les animaux

"meilleur livre 1963" (sélectionné par "Loisirs-Jeunes")

LA GÉOGRAPHIE

notre planète - ses peuples - ses ressources

L'HISTOIRE

la Civilisation depuis ses origines

chaque volume relié pleine toile (20 x 27,5 cm), sous jaquette, 368 pages, environ 2 000 illustrations en noir et en couleurs, glossaire des termes scientifiques, index.



JE SÈME A TOUT VENT